
Professor Eric J. Rubin, MD, PhD Ravikiran Raju

Where proteins go to die: Elucidating the physiological and therapeutic significance of the Clp protease complex in *Mycobacterium tuberculosis*

Abstract

Microbiologists have long focused on transcription as a main source of physiological regulation in bacterial adaptation. However, the time scale on which certain cellular responses must be coordinated dictates that a more rapid system be in place to deal with sudden environmental stresses. In eukaryotes, understanding the proteasome and ubiquitin-tagging has led to an appreciation for protein turnover as a mechanism for rapid adaptation.

Like eukaryotes, bacteria possess several proteolytic complexes that degrade proteins into smaller polypeptides and amino acids. These enzymes were discovered as maintainers of protein quality control, through recognition of aberrant protein products, but recent studies have suggested that they play an active role in regulation of cell processes through degradation of endogenous proteins. Surprisingly, a genome wide screen for essential genes in *Mycobacterium tuberculosis* (Mtb) found numerous proteases to be essential for growth, providing evidence that degradative regulation may be critical for survival. One essential complex, Clp protease, was intriguing as it appeared to have a divergent structure in Mtb, and was largely dispensable for growth in most other organisms. In order to study the importance of protein turnover and degradative regulation in Mtb, I chose to study Clp as a model.

I confirmed that Clp was required for normal growth in mycobacteria through targeted genetic engineering, and demonstrated that depletion of Clp was bactericidal. We hypothesized that a protease would be essential because it might prevent accumulation of toxic proteins or repressors of vital processes. To understand why Clp protease was so critical, I conducted proteomic analysis comparing wildtype and Clp-depleted cells to identify substrates of the protease. In line with our hypothesis, I identified WhiB1, a redox-sensitive transcriptional

repressor. Blocking degradation of WhiB1 by Clp resulted in death, suggesting that the importance of Clp can be partially explained by its action on the repressor.

Finally, taking advantage of known Clp-specific inhibitors in *S. aureus,* we showed that Clp could be targeted with small molecules in Mtb. The elucidation of novel drug targets and small molecules active against Mtb is crucial due to the overwhelming prevalence of the disease and rises in drug resistant forms.

Table of Contents

Acknowledgements

This thesis owes its existence to the cumulative effort of countless people who have invested their time and effort into ensuring my development as both a scientific thinker and a compassionate individual.

Words cannot express the gratitude I have towards my thesis mentor, Eric Rubin. Eric found me during a stage in my training when I was, for lack of a better word, lost. Disenchanted by the opportunities for conducing global health related research as a second year MD/PhD student at Baylor College of Medicine (BCM), I had contemplated giving up the PhD and focusing on clinical training. Through a series of numerous phone conversations (during which I'm sure he was roaming the halls of Armenise), Eric provided excellent and personal advice on what would be the best decision for my career. Though it might seem differently, as I now submit this thesis after being in his lab for the last three years, not once did he even suggest transferring for my PhD. In fact, I clearly remember him telling me that it would be a bad idea to forgo my MD/PhD funding at BCM and transfer to Harvard. But I was so moved by the unbiased mentorship he was willing to provide to a complete stranger that I knew that I couldn't give up the opportunity to learn even more from him. His guidance over these years, in helping me navigate both scientific and personal hurdles, has been formative and life-changing.

It makes perfect sense that such an outstanding mentor would create a one-of-a-kind lab environment. Throughout my undergraduate education and rotations at both BCM and Harvard I've worked in many research groups, but there is no place quite like the Rubin lab. Whether talking science or playing pranks (baseballs, pipettes, surprise parties just to name a few that come to mind), I couldn't possibly think of a better place to work. I entered the Rubin lab, never having done molecular biology research, and nervous that Meera Unnikrishnan (the post-doc who pioneered this project) had already left. Despite this, numerous people took me under their wing

and taught me the ropes. This thesis would have never materialized if I didn't have that initial guidance and for it I thank the "old guard" that has moved on from the Rubin lab (Mike Chao, Babak Javid, Vidhya Krishnamoorthy, Shoko Minami, Giovanna Poce, Magnus Steigedal, and Wen Zheng). A special thanks to Mike for helping with all the hurdles that a PhdD presents, from navigating the PQE, DAC meetings, and finally this dissertation itself. And of course, this project would not even exist if it were not for the initial work of Meera, so I thank her and Eric for trusting me to carry on its successes.

Every lab has that one person whose dedication and hard work keep the lab functioning. In our lab, that person is Jun-Rong Wei. Fielding endless questions every day, Jun has happily guided every single person in the lab through difficult portions of their projects. I may have taken the lions share of her time, especially early on, but the combination of her sound scientific judgment, tireless smile, and endless waving hello have kept me afloat during my time in lab. Beyond Jun, I want to thank all the current Rubin lab members (Emmy Balon, Amy Barczak, Cara Boutte, Marte Dragset, Kristi Guinn, Karen Keiser, Amanda Martinot, Alissa Myrick, Annie Park, Jessica Pinkham, Hesper Rego, Brian Schuster, Flavia Sorrentino, Andrej Trauner, and Jason Zhang) who made Armenise 439 and now HIM 1007A such an amazing place to come to work everyday. Specifically, I wanted to thank two people, Andrej and Jason, both of whom were teaching fellows with me this past semester and came to my rescue with respect to both teaching responsibilities and data analysis for this thesis, as my dissertation deadline loomed closer and closer. Andrej, I look forward to the day your multi-million dollar biotech IPOs; don't forget your old friends! And Jason, I was sad when you chose to come to Harvard instead of BCM, but reconnecting over these past years and hashing out science and life paths, has made the move to Boston that much more amazing.

By far, the most valuable lesson I've learned during my PhD training is the importance of collaboration and how working with people from disparate technical backgrounds can push

science forward by leaps and bounds. As is clearly evidenced time and time again in this thesis, my work would not have been nearly as interesting if it were not for the collaborations that we have with numerous groups. Specifically, I'd like to thank:

- **Professor Alfred Goldberg and senior scientists Tatos Akopian and Olga Kandror.** Fred's lab was like a second home for me, and Tatos and Olga taught me everything I know about protein biochemistry.
- **Professor Steven Gygi and post-doctoral fellow Mark Jedrychowski** who put up with me as I fumbled my way through learning proteomics, and graciously gave the reagents and machine time that enabled Chapter 3
- **Professor Stephan Sieber and graduate student Evelyn Zeiler** who openly shared their expertise in activity-based protein profiling, rapidly worked to make sure that the data presented in Chapter 4 could take shape before my dissertation deadline, and truly enabled the translational focus I had strived to achieve in my doctoral work.
- **Professor Sarah Fortune and lab** who through weekly group meetings helped to shape the ideas and hypotheses that formulated this project.
- **Noman Siddiqi and Larry Pipkin in the BL3 lab.** I don't understand how they suit up every single day, but am so grateful for their maintenance of our facilities.

Moving beyond research specifically, I'd like to thank the BPH program (Marianne Wessling-Resnick, Elaine Carpentier and Ruth Kenworthy) and my Dissertation Advisory Committee (Dan Finley, Sarah Fortune, and Marcia Goldberg) who have contributed to the fight of making my PhD as painless as possible. And along those lines, I couldn't be more excited about the Dissertation Defense Committee (Manoj Duraisingh, Sabine Ehrt, Sarah Fortune, and Dyann Wirth) that has come together to evaluate my doctoral work. I thank them for taking the time to make sense of this thesis, and hope they find it both enjoyable and passable!

I have been grateful throughout my life to have the guidance of many amazing mentors. The decision to leave BCM created a lot of turbulence, but I am so thankful for the support of several key BCM faculty members (Mary Estes, Frank Kretzer, Robert Parkerson, and Stephan Spann) who saw a unique opportunity for me in transferring to Harvard. Especially Frank, who is flying all the way up from Houston just for my defense! And coming back to Harvard has been great for

reconnecting with mentors from my undergraduate days, and I thank them (Steve Bloomfield, Devesh Kapur, Sean and Judy Palfrey, and Noreen Tuross) for continuing to put up with my relentless questioning of issues related to career, life, and love. By affording me the opportunity to be a resident tutor in Adams House, the Palfreys enabled the move from Houston to Boston, and have given me one of the most rewarding jobs in my life; to return the favor (of the priceless mentoring I have received) to the next generation of idealistic college students.

In closing, I'd like to thank the countless classmates and friends who have been such amazing confidants along this path. To my amazing and supportive girlfriend, Lauren, I would have gone crazy these last few months if it were not for the calmness and love that she exudes so effortlessly. And lastly, to my family, I owe so, so much. I've dedicated my entire life to the memory of my mother who through life and death taught my brother and I the importance of compassion and a life devoted to the service of others. This thesis is just a step along a career path that I hope will make her proud. To my brother, Sagar, who I am excited to embark on a life a global health work with. And to my father, thank you for dedicating your life to ensuring that Sagar and I succeed at whatever we put our minds to.

Chapter 1:

An introduction to the cell-associated, compartmentalized proteases of bacteria

This chapter has been submitted as a review article:

Raju RM, Goldberg AL, Rubin EJ. The Bacterial Proteolytic Complexes are Key Regulators of Cellular Processes and Tractable Therapeutic Targets. *Nature Reviews Drug Discovery* **submitted**

Section 1.1 – Dissertation Overview and Attributions

With the global prevalence of drug-resistant *Mycobacterium tuberculosis* (Mtb) on the rise, there is an urgent need to discover new antibiotics to aid in the fight against the deadly pathogen. An important step in the path to new antimycobacterial agents is the validation of vulnerable bacterial protein targets whose chemical modulation results in bacterial death. My doctoral work has focused on essential Clp protease in Mtb, with the dual purpose of understanding why this cytoplasmic protease is so important for cell survival and identifying small molecule modulators of Clp activity that could serve as lead compounds for antimycobacterial development.

Clp protease belongs to a broader class of cell-associated, compartmentalized proteolytic complexes that, in broad strokes, resemble the eukaryotic proteasome. These enzymes maintain protein quality control in the cell by removing abnormal protein products. They also serve a post-translational regulatory function, through the degradation of endogenous bacterial proteins. Despite our growing appreciation of the critical role that these enzymes play in both normal growth and virulence, there are currently no antibacterial protease modulators. This is especially surprising considering that proteases are frequently targeted in the treatment of certain cancers, coagulopathies, and viral illnesses.

Given the potential of this class from a therapeutic perspective, I introduce the bacterial compartmentalized proteases in Chapter One, focusing on the important physiological processes they regulate and the current repertoire of specific small molecules that modulate their activity. This tripartite introduction (characterization of the proteases, physiological importance, and chemical modulation) sets the stage for and mirrors the three following chapters, where I focus on Clp protease in Mtb.

In Chapter Two, I set out to understand the requirements of Clp protease in mycobacteria for growth in vitro and during infection. Though Clp is an extremely well-studied enzyme in model organisms such as *Escherichia coli* and *Bacillus subtilis,* a targeted study of the complex in Mtb is warranted for several reasons. First, Clp is dispensable for normal growth in these organisms, but a genome wide screen for essential genes conducted in our laboratory suggested that the protease was absolutely required for normal growth in Mtb. Second, most organisms possess one *clpP* gene, which forms the proteolytic core of the enzyme, but mycobacteria encode two homologous genes, *clpP1* and *clpP2*, situated in a single operon. Through a combination of genetic and biochemical studies, I was able to show that both *clpP1* and *clpP2* are independently essential for normal growth in *Mycobacterium smegmatis* (Msm, a non-pathogenic, fast-growing relative of Mtb that I often use as a model organism in this thesis) and that the Mtb forms of ClpP1 and ClpP2 interact together to form a single proteolytic core, denoted ClpP1P2. Taking advantage of this, we expressed active site mutants of ClpP1 in Mtb, which had a slight dominant negative effect on growth in vitro. These mutants, however, were highly attenuated in a mouse model of infection, suggesting that Clp protease has a role in infection as well.

The logical question that follows after confirming the importance of Clp protease for growth and infection is “why?”. I focus on this in Chapter Three. I hypothesized that bacteria lacking Clp protease would die because the enzyme was required to degrade a set of proteins whose accumulation resulted in toxicity or repressed normal growth. To systemically identify the set of substrates degraded by Clp, I took advantage of conditional mutants in which the levels of Clp could be modulated, and conducted quantitative, global proteomic analysis comparing the proteomes of wildtype and Clp-depleted mycobacteria. This analysis revealed numerous proteins that were over-represented in mutant bacteria. After confirming that over-representation was due to lack of Clp-dependent proteolysis and not due to a non-specific stress response resulting from depletion of Clp, I identified a transcriptional repressor, WhiB1, as a putative Clp substrate. Interestingly, blocking the Clp recognition motif on WhiB1 resulted in an ultra-stable allele that

was toxic. As this altered allele of WhiB1 was still functional, it suggested that dysregulation of WhiB1 proteolysis results in WhiB1-dependent toxicity. While there are likely many important Clp substrates, these studies illustrate that the toxicity of Clp depletion can be partially explained by the stabilization of the transcriptional repressor, WhiB1, which may inhibit cellular processes required for normal growth.

In addition to shedding light on important biology, the study of essential bacterial enzymes is also attractive for the development of new antibiotics. As genetic depletion of Clp protease in mycobacteria was rapidly bactericidal, Chapter Four addresses the hypothesis that small molecule inhibition would also be cidal. In order to chemically validate Clp protease as a drug target, I took advantage of studies published on trans β-lactones as Clp-specific inhibitors in *Staphyloccus aureus* and *Listeria monocytogenes*. In these organisms, Clp protease is not essential for growth in vitro, but is required during infection. Accordingly, the β-lactones are not lethal in *S. aureus* or *L. monocytogenes,* but do abolish virulence. Given the divergent essentiality of Clp, however, I hypothesized that the β-lactones could act as growth inhibitors in mycobacteria. Accordingly, we identified a particular β-lactone, EZ120, that was bactericidal in both Mtb and Msm. The compound inhibited ClpP1P2 proteolysis in an in vitro purified protein peptidase assay, and by conducting activity-based protein profiling studies using derivitized lactones and intact mycobacterial cells, we were also able to show that the β-lactones interacted with Clp protease in vivo. Together, these studies demonstrate that the β-lactones inhibit Clp protease in mycobacteria, and that targeting Clp protease is an ideal strategy for the development of new antimycobacterial compounds.

In Chapter Five, I conclude by summarizing the significance of these findings, and proposing further studies that would increase our understanding of how Clp protease and the other essential proteolytic complexes in mycobacteria regulate important biological processes. In my opinion, we are on the cusp of greatly enhancing our understanding of how bacteria adapt to new

environments (including survival in the host during infection) by taking into account the rapid adaptation that cells initiate through targeted proteolysis. It is also critical that, as an understanding of these important enzymes emerges, we exploit this knowledge towards the development of an increased antibiotic arsenal. As I hope to convey in this thesis, the study of the cell-associated, compartmentalized proteases is one of the key areas of microbiology where the elucidation of novel biology can go hand in hand with translational medicine.

Attributions. The manuscript that comprises Chapter One was researched and written by myself, with valuable editorial input from Professors Alfred Goldberg and Eric Rubin.

Section 1.2 – The bacterial proteolytic complexes are key regulators of cellular processes and tractable therapeutic targets

Ravikiran M. Raju[a], Alfred L. Goldberg[b], Eric J. Rubin[a,1]

[a]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02138 USA

[b] Department of Cell Biology, Harvard Medical School, Boston, MA 02138 USA

[1] Corresponding author: erubin@hsph.harvard.edu, ph: (617) 432-3337

INTRODUCTION

In recent years, the dramatic rise of drug-resistant bacterial species has led organizations such as the World Health Organization to warn of a post-antibiotic era, in which the current antimicrobial regimens would be largely ineffective against infectious pathogens. In fact, well-established resistant strains such as methicillin-resistant *Staphylococcus aureus* (MRSA) and extremely-drug resistant *Mycobacterium tuberculosis* (XDR-TB) have already begun to erode

earlier advances in treatment and control. As resistance has now been documented in nearly every human pathogen, there is an urgent need to discover new antibiotics. The simplest way to circumvent resistance is to develop drugs with novel mechanisms of action, which will require a new set of antibacterial targets.

Proteases, enzymes that hydrolyze the peptide bond between amino acids, are an attractive class of targets for the development of therapeutics. These enzymes are highly susceptible to inhibition with pharmacologically attractive compounds. In fact, proteases are targeted in the treatment of numerous diseases. Examples abound for easily accessible extracellular proteases, which function in blood pressure regulation and coagulopathies, and intracellular proteases as encoded by HIV and other viruses. Most recently approved for human use is boceprevir, an inhibitor of the Hepatits C virus (HCV) NS3/4A protease, which cleaves the HCV polyprotein into the mature proteins that allow viral packaging and replication[1]. Currently, it is estimated that 5-10% of all pharmaceutical targets being pursued are proteases[2]. However, despite the pharmaceutical successes with protease modulators, there are currently no protease inhibitors used to treat bacterial infections.

As in eukaryotes, bacterial proteases come in many flavors. They can be broadly separated into two major classes: proteolytic complexes that completely degrade proteins into amino acids and single subunit proteases that engage in site specific cleave of proteins into smaller protein products. These classes can be further divided based on their location, either intracellular or secreted. Significant attention has been given by biologists to the secreted proteases in pathogenic bacteria, due to their role in virulence. The majority of them are unique to specific pathogens and are single subunit enzymes that target a specific protein or set of substrates in an unregulated fashion. For example, organisms that thrive on mucosal surfaces, such as *Neisseria gonorrhoeae* and *Streptococcus pneumoniae*, secrete IgA1 proteases, which inactivate IgA and impair immune defense mechanisms at sites of infection[3]. As in the case of IgA1 protease, host

proteome modulation is a tool employed by other bacteria to promote disease. While the effect of most single subunit proteases is determined by their cleavage specificity of particular sequences, proteases involved in intracellular protein turnover generally have a broad substrate range and must therefore be tightly regulated to avoid excessive or non-selective proteolysis of endogenous bacterial proteins. There are several conserved bacterial proteolytic complexes, each tightly regulated by two key mechanisms. First, these enzymes are all large barrel-shaped, multimeric, compartmentalized complexes, in which the proteolytic active sites are isolated in a chamber away from the cellular environment. Protein substrates are taken up into these degradative chambers and hydrolyzed to small peptides in a highly processive manner. Second, each of these proteases contains regulatory components, either separate domains or proteins, that control accessibility to the active sites and specificity for particular proteins. Thus, despite carrying out a destructive process, these tightly regulated enzymes are able to provide an efficient mechanism for maintaining protein quality control by degrading misfolded polypeptides, and coordinating post-translational regulation through targeted degradation of endogenous proteins.

Neglected from a drug development perspective until recently, the bacterial degradative complexes are attractive targets for several reasons. First, they are widely conserved among bacteria, thus increasing the chance of broad-spectrum coverage of any small molecule modulators. Second, these enzymes perform numerous essential functions in the cell. Degradative proteases were initially identified in *Escherichia coli* for their role in general quality control of the proteome, via the degradation of abnormal proteins, but were largely found to be non-essential. More recent studies in pathogenic bacteria have revealed that these proteases play numerous critical, regulatory roles in many physiological processes, and are in almost every case indispensable for either normal growth or virulence. For example, while Clp protease is non-essential for vegetative growth in *E. coli*, it was found to be essential both in vitro and during infection in *Mycobacterium tuberculosis*[4]. Third, as these enzymes are large complexes with regulatory and catalytic components, numerous surfaces could be effectively targeted by small

molecules. Much work has gone into elucidating the biochemistry and structural biology of these key enzymes (at least in *E. coli*, although not in pathogens), which should facilitate target-based drug development approaches.

We begin with a brief description of the general properties of these enzymes, and an introduction to the specific conserved, proteolytic complexes found in bacteria. Next, we validate a target-based approach to drug development by discussing the physiological significance of these enzymes and highlighting key processes regulated by the bacterial proteolytic complexes. As the study of these enzymes has been the focus of many talented researchers, it is impossible to present a complete overview here. Instead, we aim to illustrate over-arching themes and concepts that facilitate a broad understanding of the degradative proteases and how they coordinate important cellular processes. Finally, we close with an in-depth look at the recent efforts to develop specific modulators of the degradative proteases and the lead compounds that could provide a direct path to a novel class of antibiotics.

THE KEY PROKARYOTIC DEGRADATIVE PROTEASES

Bacteria contain a key set of conserved, cell-associated degradative proteases that are responsible for turnover of intracellular proteins. The main proteases include the Clp, FtsH, Lon, HslUV, the prokaryotic proteasome, and the HtrA family of enzymes. These complexes are found in every cell compartment: in the cytoplasm, on the cell membrane, and in the periplasm. With the exception of the bacterial proteasome, they are ubiquitous within the Eubacteria kingdom and are even conserved in the chloroplast and mitochondria of eukaryotes. Each protease can be broadly characterized as having two major components: a recognition domain that is responsible for initial binding to a protein substrate, and a proteolytic domain that carries out peptide bond hydrolysis. With the exception of the periplasmic HtrA family of proteases, these enzymes also possess a hexameric ATPase subcomplex, which provides energy for substrate unfolding, translocation into

the proteolytic domain, and also regulates the proteolytic compartment[5]. All these enzymes are hexameric or heptameric complexes that are much larger (between 500,000 and 1,000,000 kDa) and complex than the typical extracellular and single subunit intracellular proteases (20,000 to 30,000 kDa). Lon, FtsH, and HtrA proteases encode both the recognition domain and proteolytic domain in polypeptide, while in Clp, HslUV, and the proteasome the recognition and proteolytic functions are separated on two separate protein complexes **(Table 1.1)**. The proteases that exhibit ATPase activity fall into the broad class of AAA+ enzymes, or ATPases associated with diverse cellular activities, which have been well-characterized and studied in numerous organisms[6]. Though not covered in this review, these ATPase complexes have also been shown to act as independent chaperones, aiding in the remodeling of particular proteins[7].

Regardless of the structure of these domains, and unlike the majority of proteases which are single subunit enzymes, each of the degradative enzymes requires multimerization of subunits to form a barrel-shaped structure with the active sites of each proteolytic subunit facing inwards into a hollow chamber. This protected structure ensures efficient degradation of substrates delivered by the recognition domains, and prevents non-specific hydrolysis of random proteins. Typically, continued ATP hydrolysis and/or substrate binding catalyzes a conformational change of the complex that enables access of substrates to the hydrolytic active sites in the degradative chamber.

Substrates are targeted for degradation by several mechanisms. All degradative proteases rapidly degrade proteins that lack well-defined tertiary structures. Some substrates possess a small, unfolded stretch of amino acids, which serves as a motif that interacts with the protease recognition domain and leads to targeting for degradation[8]. In other cases, substrates require modification to unmask a hidden sequence that is recognized by the protease. For example, degradation motifs may be revealed through phosphorylation that causes a reorganization of exposed residues, initial cleavage by a site-specific protease, or loss of protein-protein

interactions that normally protect a specific protease recognition sequence[9]. Other substrates require adapter proteins to target them to a protease for degradation[10].

Table 1.1 - Classification and features of the bacterial degradative proteases

	Clp	Lon	FtsH	HslUV	Proteasome	HtrA
Catalytic mechanism	Serine protease (Ser-His-Asp catalytic triad)	Serine protease (Se-Lys catalytic dyad)	Zn^{2+} metallo-protease	N-terminal threonine	N-terminal threonine	Serine protease (Ser-His-Asp catalytic triad)
Cellular location	Cytoplasmic	Cytoplasmic	Membrane-bound	Cytoplasmic	Cytoplasmic	Periplasmic
ATPase activity?	Yes	Yes	Yes	Yes	Yes	No
ATPase on separate protein?	Yes (ClpA/C Gram neg., ClpX/C Gram pos.)	No	No	Yes (HslU)	Yes (Mpa)	N/A
Examples of additional adapters	ClpS RssB MecA UmuD SspB	PolyP PinA DNA	HflK MgtR HflC SpoVM HflD	None reported	PafA	CpxP
Specific recognition motifs*	SsrA tag N-end rule Endogenous motifs	SsrA tag Endogenous motifs Linear, aromatic residues	Endogenous motifs	Endogenous motifs	Pupylated substrates	Endogenous motifs Linear, aromatic residues
Overview of Essentiality**						
Gram (+)	Essential for virulence	N/A	Not investigated	Not investigated	N/A	Essential for virulence
Gram (-)	Essential for virulence	Essential for virulence (over-expression lethal in *E. coli)*	Essential for growth	Not investigated	N/A	Essential for virulence
Mtb	Essential for growth	N/A	Essential for growth	N/A	Essential for virulence	Essential for growth

*** All proteases are capable of recognizing unfolded proteins**
****Overview is based on existence of at least one study in a Gram (+) or Gram (-) bacteria, or Mtb. N/A denotes absence of protease, in that particular class of organisms.**

Clp Protease

Clp is a serine protease with a characteristic Ser-His-Asp catalytic triad. The active complex consists of individual ClpP proteolytic subunits stacked into two heptameric rings. Specificity for protein substrates is determined through separate ATPase complexes, hexameric rings that interact with the apical surfaces of the ClpP tetradecamer. In Gram positive bacteria, either of two ATPase complexes, ClpX and ClpC, can interact with ClpP to form the active ClpXP and ClpCP proteases, while in Gram-negative bacteria, ClpA and ClpX cap the proteolytic ClpP tetradecamer[11]. These ATPases can recognize a diverse set of motifs present almost anywhere within the protein; N-terminal, internal and C-terminal recognition motifs have been described. For instance, the "N-end rule" in prokaryotes states that the presence of an abnormal amino-terminal residue can lead to rapid degradation[12]. The adapter ClpS assists in the binding of polypeptides with large, hydrophobic N-terminal Phe, Leu, Trp or Tyr, and delivering proteins with these destabilizing N-terminal residues to ClpAP or ClpCP for degradation[13]. While the hexameric ATPases are capable of recognizing various proteins, additional adapters such as ClpS also interact with the ATPases to facilitate substrate degradation.

FtsH Protease

FtsH (also referred to as HflB in some organisms) is a homohexameric complex of Zn^{2+} metalloprotease subunits. With two N-terminal transmembrane domains, the C-terminal proteolytic machinery of FtsH lies in the cytosol, and thereby enables degradation of both cytoplasmic and membrane-associated proteins[14,15]. Unlike Clp, both ATPase and proteolytic function are encoded on the same polypeptide, and upon multimerization, ATPase domains extend outwards while the proteolytic machinery is protected and buried internally. FtsH is thought to recognize unfolded stretches of amino acids as well as specific recognition motifs present at either the ends or internally within a protein. Between the two transmembrane domains of FtsH, a periplasmic stretch coordinates interactions with the extracytoplasmic adapter complex HflKC, which inhibits degradation of membrane-associated proteins, thus favoring recognition of

cytoplasmic substrates[16]. In the cytoplasm, other adapters such as HflD and MgtR interact with the protease to enable degradation of specific proteins.

Lon Protease

Lon protease is a serine protease that contains a Ser-Lys catalytic dyad. The single polypeptide that multimerizes into the active hexameric complex contains the N-terminal substrate binding and oligomerization domain, the ATPase domain, the substrate sensor and discriminatory domain (SSD), and the C-terminal proteolytic domain[17]. Substrates are recognized by the substrate binding domain, then unfolded via ATPase activity, and finally threaded into the proteolytic chamber. As with Clp protease, Lon appears to recognize an extremely diverse set of motifs in proteins. Additionally, Lon interacts with small molecule adapters such as polyphosphate or chromosomal DNA to coordinate degradation of specific proteins during starvation when the intracellular polyphosphate concentrations are elevated and upon stresses that cause DNA damage[18].

HslUV Protease and the Prokaryotic Proteasome

The HslUV protease and the proteasome comprise a unique family of cytoplasmic proteases that contain an N-terminal threonine residue responsible for hydrolysis. Lacking a canonical catalytic dyad or triad, these enzymes exhibit a unique mechanism of catalysis in which the threonine hydroxyl acts as the nucleophile to hydrolyze peptide bonds, while the N-terminal amino group acts as a general base to facilitate catalysis[19]. The well-conserved HslUV protease (sometimes referred to as ClpYQ) resembles the ClpCP and ClpXP proteases in overall structure. While the HslU subunit is a member of the AAA+ ATPases and has significant homology to ClpX, the HslV subunit shares a high-degree of similarity with the proteolytic beta-subunits of the eukaryotic 20s proteasome[20]. Substrate recognition is afforded through the HslU ATPase which also provides the energy for unfolding and translocation.

The prokaryotic proteasome, however, is only present in a few pathogenic bacteria, including mycobacteria, nocardia, and *Propionibacterium acnes.* The bacterial proteasome is composed of four, stacked proteolytic rings and closely resembles the eukaryotic 20s proteasome with two inner hepatameric rings composed of PrcB (beta subunits) and two outer heptameric rings of PrcA (alpha subunits)[21]. However, while eukaryotes encode seven different but homologous alpha and beta subunits, the prokaryotic proteasome only contains one of each type of subunit. In addition, bacteria encode a proteasome regulatory ATPase hexamer, Mpa, but interaction with the 20s component has yet to be confirmed[22]. *M. tuberculosis* can conjugate a marking protein (Pup) to substrates leading to their recognition by Mpa and degradation by the prokaryotic proteasome[23]. Pup thus functions in a homologous manner to ubiquitin, although the structure and chemistry are divergent from eukaryotic ubiquitin.

HtrA Protease

The HtrA family (also known as Deg family) of serine proteases are different from the enzymes above in several key respects. First, HtrA proteases are secreted into the periplasm and are, therefore, responsible for degradation of outer membrane, periplasmic, and foreign proteins. Second, they lack an ATPase domain and do not require ATP for substrate hydrolysis. Third, they exhibit significant chaperone activity that is independent from their role as proteases. This chaperone activity appears to be crucial for the proper folding and localization of key outer membrane proteins[24]. Fourth, despite a funnel-shaped structure which also leads to compartmentalization of active sites, instances of limited processivity by HtrA have been documented[25].

Some bacteria contain two to three different HtrA-like proteins. All HtrA proteins share a similar domain structure, with an N-terminal protease domain and a C-terminal PDZ domain, which is broadly involved in protein-protein interactions and substrate recognition. In *E. coli*, which contains three HtrA proteins, DegS is a funnel-shaped trimer with protease domains forming a

central core, and the PDZ domains protruding outwards[26]. These protrusions play a critical role in substrate recognition. DegP is found as a trimer in its inactive state, but upon substrate binding to the PDZ domain, the trimer polymerizes into higher-order structures such as 12mers, 15mers or even 18mers[27]. In each of these arrangements, the protease domains always form a central chamber, while the PDZ recognition domains protrude outwards. DegQ, the third Htra-like protease in *E. coli*, is similar in structure to DegP.

PHYSIOLOGICAL ROLES OF THE DEGRADATIVE PROTEASES

One of the most critical components of target-based approaches to antibiotic development is to demonstrate the essentiality of a particular target. If genetic studies reveal that a protein of interest is required for the growth of a pathogen, either in vitro or in vivo, there is a strong likelihood that chemical inhibition of that target will result in bacterial death during infection. Despite the conservation of the degradative proteolytic complexes, no generalizable conclusions can explain their essentiality or conditional requirements, which seems to vary across species. For example, Clp protease is dispensable for in vitro growth of most pathogens, but is essential for virulence and in vivo survival. In *M. tuberculosis,* however, Clp is absolutely required for in vitro and in vivo growth in[28]. FtsH protease appears to be essential for normal growth in *E. coli,* while Lon only becomes essential during certain types of stress[29]. This diverse pattern highlights that though the proteases may be conserved across Eubacteria, they have evolved divergent functions. Nonetheless, all of these enzymes are largely essential for growth during infection, thus validating the proteases as broad targets for the development of anti-virulence agents, and in the case of some species, traditional growth inhibitors.

Recent studies have begun to paint a rich portrait of how the degradative proteases are able to regulate numerous cellular processes, through targeted degradation of specific regulatory proteins. Though no unifying theme on the functional significance of these proteases has

emerged to date, there are several key aspects of physiology that they coordinate. These broad areas include general protein quality control, specific stress responses, cellular replication and differentiation, and virulence.

General protein quality control

It was the initial attempts to understand how cells eliminate abnormal proteins that led to the discovery of ATP-dependent protein degradation and the characterization of the proteolytic complexes, Lon and Clp. Protein abnormalities can arise in a variety of ways, including mutations in DNA, errors in transcription or translation, post-translational damage (ie. by heat or oxygen radicals), and spontaneous denaturation. All such changes can produce potentially toxic proteins, and the degradative proteases minimize the impact of this toxicity through three main mechanisms.

First, all of the degradative proteases have an affinity for abnormally folded proteins. Generally, proteins are globular and hydrophobic regions are internal, but mutations and damage lead to loose, unfolded structures with exposed hydrophobic residues. As mislocalization and misfolding can occur in all cellular compartments, each space is equipped with general quality control enzymes. In the periplasm, HtrA is the predominant protease responsible for clearance of misfolded proteins. The PDZ domain of HtrA proteases binds unfolded stretches in misfolded or mislocalized proteins and directs them towards the proteolytic chamber for degradation[30]. In the cell membrane, the anchored FtsH insures protein quality control by degrading proteins that are misfolded or unassociated with other interacting partners. For example, individual components of the F1 ATPase, involved in generating the proton motive force, and the Sec secretion system that are inserted into the membrane without the full complex are targeted for degradation by FtsH[15,31]. In the cytoplasm, Lon appears to be the primary protease responsible for degradation of abnormal proteins. The N-terminal substrate binding domain has an affinity for linear arrangements of aromatic amino acid residues, hydrophobic residues that are normally buried

internally within a protein, and the carbonyl derivatives that can form upon reaction with reactive oxygen species[8].

Second, several proteases are involved in the ribosome rescue pathway, or trans-translation, which requires degradation of polypeptides that have stalled on the ribosome during translation. During translation, ribosomal stalling may occur upon truncation or oxidation of mRNA or a depletion of required tRNAs. In response, the mRNA can be displaced by a unique species, the transfer messenger RNA (tmRNA), which directs the addition of a C terminal peptide, the SsrA sequence[32]. The modified polypeptide, once released from the ribosome, can be recognized and degraded by multiple proteases, primarily ClpXP, but also Lon and FtsH. Interestingly, recent work suggests that tmRNA pathway in *M. tuberculosis* is inhibited by the antituberculous drug pyrazinamide, which binds to and interferes with the function of ribosomal protein S1[33].

Third, proteases themselves are upregulated during conditions that promote protein misfolding such as heat shock. In fact, Lon, Clp and HslUV were initially identified as heat shock proteins as they were induced during stress and promoted the degradation of misfolded proteins. In the cytoplasm, these proteins are induced together with other major heat shock proteins, such as DnaK, GroEL, and GroES, which function as molecular chaperones to promote unfolding and refolding of damaged proteins[34]. If chaperones fail to refold their substrates, these aberrant proteins can be targeted to the cytoplasmic proteases for degradation. In the periplasm, bacteria induce the unfolded protein stress response in response to misfolded polypeptides, and this pathway is critically regulated by and dependent on the degradative proteases. In the periplasm in *E. coli*, CpxP normally binds to and inhibits the transmembrane sensor kinase CpxA. In the presence of abnormal proteins, CpxP binds to them and targets them to DegP for degradation[35]. CpxA is then free to phosphorylate the cytoplasmic response regulator CpxR[36]. In the cell envelope, RseA, a membrane-bound anti-sigma factor, binds to and sequesters RNA polymerase σE (RpoE) at the cell membrane. In the presence of misfolded proteins, DegS is activated to

cleave RseA[37]. Together RpoE, which is now free in the cytoplasm, and phosphorylated CpxR induce the expression of genes that serve to maintain the cell envelope during stress[38]. As infection of the mammalian host is highly stressful, RpoE and CpxR have been found to be essential for the viability and virulence of numerous organisms including *Salmonella enterica serovar typhimurium,* and *M. tuberculosis*[39,40].

Coordination of stress responses

In addition to coordinating the clearance of abnormally folded proteins and the cellular response to their presence, the degradative proteases are also involved in the regulation of numerous other stress responses. In stressful environments, bacteria must activate a diverse set of transcriptional responses. Each stress response can be characterized as having three phases: induction, in which transcriptional activation of effector proteins occurs, maintenance, during which key players coordinate adaptation to the stressor, and termination, which occurs after the stress has been relieved and cells must return to their normal physiological state. At each of these points, effective regulation of many stress responses is critically dependent on the degradation of particular proteins. These mechanisms have been most extensively studied in *E. coli*.

The multiple levels at which proteolytic regulation modulates stress response induction in *E. coli* is best illustrated through the master regulator of the general stress response, RpoS **(Figure 1.1)**. RpoS is a transcription factor critical for the induction of numerous pathways that promote stress adaptation including genes involved in heat shock, anaerobic metabolism, and DNA repair[41]. RpoS is transcribed and translated normally, but its level in cells remains low due to efficient degradation by ClpXP protease[42]. However, entrance into stationary phase or the presence of various stress-inducing signals such as UV irradiation, heat, or acid, results in a dramatic stabilization of RpoS. Proteolysis requires the adapter protein RssB which, when phosphorylated, binds RpoS and recruits the transcription factor to ClpXP[43]. During stress induction, RssB is

sequestered by several different inducible anti-adapters, generally named Ira proteins, thus stabilizing RpoS[44].

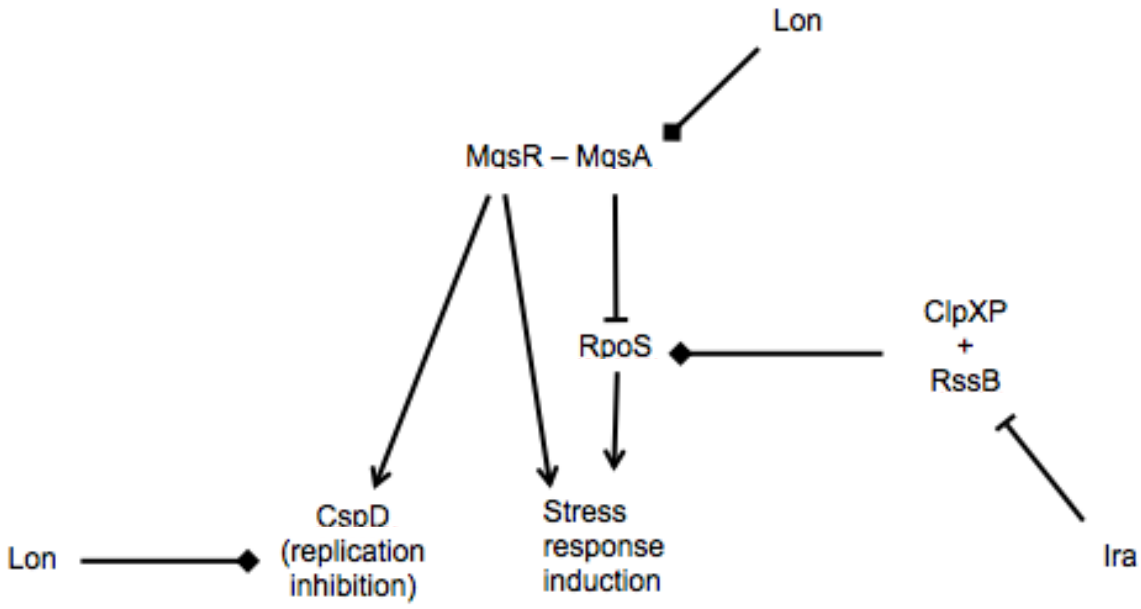


Figure 1.1 Role of proteolysis in regulation of RpoS
Levels of RpoS are regulated by several proteolytic circuits (degradation denoted by diamond arrow head). During normal, vegetative growth, RpoS levels are kept low by constitutive ClpXP degradation. Proteolysis requires the adapter protein RssB, which binds to RpoS and promotes recognition by ClpXP. During certain stresses, a set of Ira proteins are produced that bind and sequester RssB, thus stabilizing RpoS and allowing induction of stress adaptation. In an additional layer of control, the antitoxin MqsA inhibits transcription of RpoS during normal growth. Stressful environments favor degradation of MqsA by Lon protease, and subsequent activation of the cognate toxin MqsR, which induces the transcription of stress response loci including genes that promote inhibition of replication, such as CspD. Furthermore, CspD is also degraded by Lon protease, and this breakdown of CspD is required for cells to resume normal growth after surviving a stressful insult.

An additional layer of protease-mediated regulation of RpoS involves the MqsAR toxin-antitoxin (TA) module. Briefly, TA systems usually consist of two components: a stable, protein toxin that effects cell growth via a variety of mechanisms, and a rapidly degraded antitoxin, either a protein or RNA, that inhibts toxin activity. Under normal circumstances, both the toxin and antitoxin are co-transcribed. However, when transcription of the locus is inhibited, the unstable antitoxin is titrated away, most often through degradation, and the stable toxin is left uninhibited. Most commonly, toxins exhibit their toxicity through endonuclease activity. In some cases, such as for the MqsAR locus, both toxin and antitoxin can also act as transcription factors. In the absence of any stress, the antitoxin MqsA actively represses *rpoS* transcription[45]. However upon the induction of oxidative stress, Lon protease rapidly degrades MqsA. In addition to de-repression of the rpoS locus, the now-released MqsR toxin induces the transcription of stress response loci including a gene that promotes inhibition of replication, CspD[46]. CspD is also proteolytically regulated by Lon protease, and this breakdown of CspD is required for cells to resume normal growth after surviving a stressful insult[47]. Strikingly, the same protease that is involved in the induction of MqsR toxicity through degradation of the MqsA antitoxin, also attenuates the MqsR response through degradation of effector proteins. Polyphosphate, a Lon protease cofactor, might play a role in this process, favoring degradation of MqsA only in conditions of energy depletion when the intracellular concentration of polyphosphate increases.

The degradative proteases are also important effectors during the maintenance phase of adaptation. In the face of altered nutrient availability, bacteria must have a rapid way to modulate metabolic activity and biosynthetic clusters. Degradative proteases are well-equipped to initiate rapid changes in the nutritional programming of a cell, by efficiently clearing away unneeded proteins, or degrading transcriptional regulators of key metabolic pathways. For example, in *Bacillus subtilis,* ClpCP down regulates central metabolic pathways during glucose starvation through degradation of GlmS (cell wall biosynthesis), IlvB (isoleucine biosynthesis), PurF (purine biosynthesis), and PyrB (pyrimidine biosynthesis)[48]. In addition, during carbon limitation, bacteria

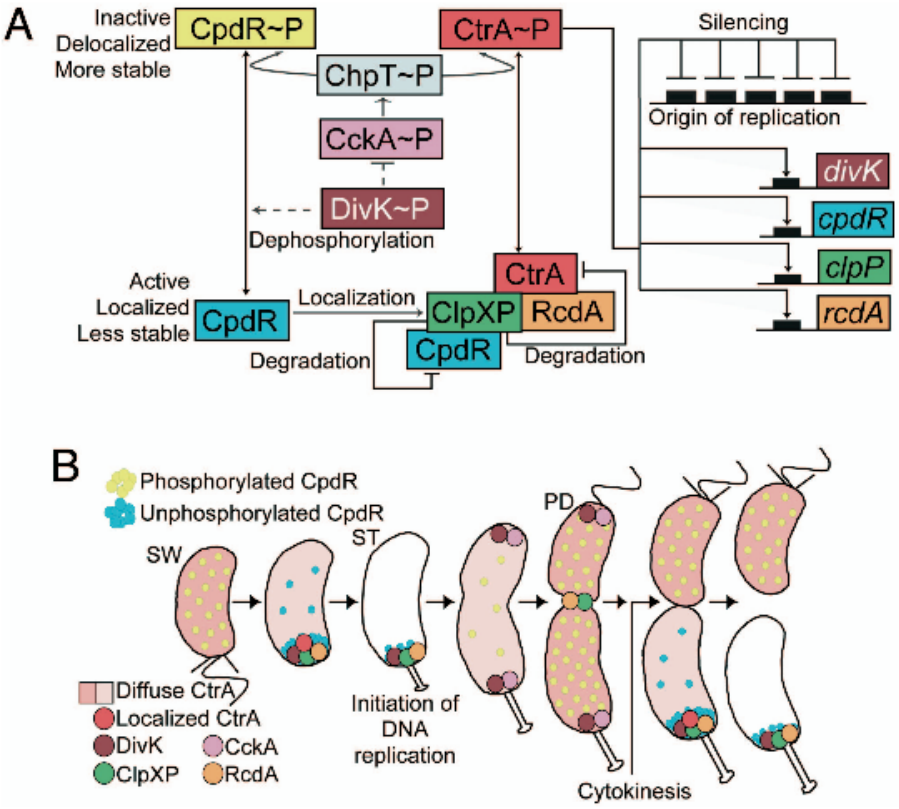
also modulate ribosomal levels to conserve energy. In *E. coli*, Lon protease is responsible for targeted degradation of the ribosomal proteins S2, L9, and L13 during amino acid starvation[49]. This conditional regulation is facilitated by the fact that Lon protease only recognizes these ribosomal proteins when bound to the co-factor polyphosphate, which accumulates in response to starvation.

Proteases are also important effectors as cells emerge from nutrient limiting conditions. ClpXP regulates the levels of Dps, a DNA binding protein that binds the chromosome non-specifically and protects the genetic material from damage while simultaneously preventing replication and transcription[50]. Moreover, ClpXP is required to degrade proteins involved in anaerobic and alternative carbon metabolism such as malate synthase and isocitrate lyase, which may facilitate adaptation once cells are shifted from a carbon-starved to a rich growth medium[51].

During the termination of stress responses, proteases can ensure rapid inhibition through the dual recognition of both transcriptional activators and effector proteins. For example, in *E. coli,* FtsH degrades sigma32, the key activator of the heat shock stress response and SoxS, a transcriptional activator in the oxidative stress response, to prevent aberrant induction of these pathways[52,53]. In *M. tuberculosis* there is an additional level of control. A down-stream effector of SoxS, Superoxide disumutase A (SodA) is pupylated and targeted for degradation by the mycobacterial proteasome[54]. Similar degradation of effector proteins also occurs in the case of DNA repair enzymes that are induced by stress. For example, the nucleotide excision repair (NER) machinery must be exquisitely controlled to prevent random mutagenesis. As NER is specifically equipped to remove thymidine dimers that result from UV damage of DNA, ClpXP actively controls the amount of UvrA, an essential protein in the NER complex present based on the frequency of thymidine dimers present in DNA[55]. Similarly, UmuD', a component of the error-prone DNA polymerase PolV complex, is degraded by ClpXP, preventing excessive chromosomal mutations after DNA has been repaired[56].

Regulation of cellular replication and differentiation

Efficient cellular replication requires the temporal coordination of several major steps including initiation of DNA replication, and cytokinesis or separation of daughter cells into progeny. Degradative proteases play a role in both DNA replication and in daughter cell division. In *Caulobacter crescentus,* the initiation of chromosomal replication is proteolytically controlled **(Figure 1.2)**. *Caulobacter* has a complex life cycle with two morphologically distinct cell types, a non-replicating motile, swarmer cell type and a replicating, stalked type. The switch between morphotypes and the initiation of DNA synthesis are both inhibited by a DNA-binding protein, CtrA. When phosphorylated, CtrA is a stable protein that binds tightly to the chromosomal origin, inhibiting DNA replication, and acts as a transcriptional factor that maintains the non-replicating, motile state[57]. Upon dephosphorylation, CtrA can be released from DNA and degraded by ClpXP via a dedicated adapter protein, RcdA, which allows replication to commence. Interestingly, at the time of the transition, CtrA degradation is localized at the nascent pole by ClpXP interactions with an anchoring protein, CpdR[58]. The Clp system is essential in Caulobacter as loss of ClpXP prevents phase transition and DNA replication.



From Iniesta et al, 2008 [58]

Figure 1.2 Proteolytic regulation of chromosomal replication in *Caulobacter crescentus*

Model for CpdR and CtrA regulation and cell cycle progression.

(A) Continuous and dashed lines indicate direct or possible indirect effects, respectively. The black boxes represent CtrA binding sites. (B) The differential spatial distribution of multiple regulatory elements that drive cell-cycle progression are shown. PD, predivisional cell; SW, swarmer cell; ST, stalked cell.

In order to divide, cells must first eliminate the division machinery that is left over at the most recently created pole. In *E. coli,* FtsZ, a tubulin homolog forms a circumferential ring around the cell and is responsible for recruiting the divisome complex of proteins and providing the contractile force required for cytokinesis. ClpXP preferentially degrades polymerized FtsZ, allowing disassembly at the site of cytokinesis after separation and reassembly at the midcell of daughter cells that are priming for a new round of division[59]. In the Gram positive organisms *B. subtilis* and *M. tuberculosis* ClpXP might have a different role, acting as an inhibitor of FtsZ polymerization[60]. In addition to directly degrading FtsZ, degradative proteases play a role in modulating the activity of proteins that interact with the divisome. For example, in *E. coli* , SulA, a replication inhibitor protein induced during the bacterial SOS-response, binds directly to FtsZ inhibiting polymerization and contractile ring formation. Upon the resolution of the SOS response, both Lon and HslUV protease rapidly degrade SulA, thus allowing the resumption of cellular division[61].

In addition to replication, population-level survival also requires differentiation of cells to survive longer times in harsh environments. Key examples of cellular differentiation and reprogramming include sporulation and biofilm formation. While many spore-forming bacteria, such as *Bacillus anthracis, Clostridium dificile, Clostridium botulinum,* and *Clostridium tetani,* are important pathogens, sporulation has been best studied in *Bacillus subtilis.* Sporulation in *B. subtilis* is regulated by controlled degradation in several respects. In the pre-divisional cell, Spo0A is the master regulator of sporulation that is activated in response to nutrient limitation and induces asymmetrical division. As phosphorylation of Spo0A is required for full activity, phosphatases such as Spo0E act as inhibitors of Spo0A and sporulation[62]. It was recently found that FtsH degrades Spo0E both in vitro and in vivo, and that accordingly, FtsH mutants were deficient in their ability to sporulate[63]. SpoVM, a small polypeptide with slow degradation kinetics, is thought to act as a decoy substrate and an inhibitor of FtsH-dependent Spo0E degradation[64]. Moreover,

both Lon and ClpCP protease degrade Spo0H another transcription factor important in initiating sporulation[65].

Another type of differentiation, biofilm formation, is critical for attachment of some pathogens in the host and for survival in the face of a host immune response and antibiotic treatment. Recently, Clp protease has been shown to enhance biofilm production of *Staphyloccus epidermidis,* a common cause of prosthetic device infections, by degrading Spx, a negative regulator of biofilms[65,66]. As Spx is conserved in related pathogens, such as *Staphylococcus aureus* and *Streptococcus pneumoniae*, Clp may play an important role in these pathogenic biofilm-forming bacteria.

Regulation of Virulence

In almost every pathogenic bacterial species that has been examined, degradative proteases have been found to act as master regulators of critical aspects of virulence, including initial colonization and prevention of killing by host responses. Establishment of infection often involves upregulation of cellular processes that facilitate adherence, invasion, and survival inside the harsh host environment. By regulating the expression of proteins present at the outer surface of the bacteria, the HtrA proteases are critical for these initial events in numerous gut pathogens. In *Campylobacter jejuni*, a common cause of gastroenteritis, HtrA stimulates adherence to the gastrointestinal epithelium[67]. In another gut pathogen, *Salmonella typhi* HtrA mutants exhibited poor systemic spread after oral inoculation in a mouse model of infection. As spread of the mutant was not affected when mice were infected intravenously, HtrA was implicated in the initial penetration of the intestinal barrier during systemic spread of Salmonella[68].

In intracellular organisms, successful infection requires adaptation to the harsh intracellular environment of host cells. One protein, MgtC, is conserved in a number of intracellular pathogens that infect macrophages, including *Salmonella typhimurium, M. tuberculosis,* and *Yersinia pestis*.

In each organism, MgtC is required for growth in macrophages, presumably due to its role in magnesium uptake[69]. In *Salmonella enterica*, MgtC is not present in bacteria growing *in vitro* despite high levels of mgtC transcript. This appears to be due to FtsH-mediated degradation, an event that likely ensures proper timing of the induction of virulence and invasion pathways. Degradation requires a short peptide, MgtR that acts as an FtsH adapter[70]. In other cases of host adaptation, the substrates are not so clear. For example, *M. tuberculosis* requires the prokaryotic proteasome for survival in the host, but it has been suggested that proteolytic activity of the complex may not be required for this survival effect[71]. Additionally, Clp protease is essential for normal growth in vitro and during infection in *M. tuberculosis*, but for unknown reasons[28].

Several Gram negative pathogens utilize a specialized protein secretory pathway, the type III secretion system, to transport proteins to host cells and modulate host processes to facilitate infection. Both Clp and Lon protease have been shown to have a role in the regulation of type III secretion (TTSS) in various organisms. Interestingly, Lon protease appears to have distinct effects on secretion in different bacterial species. In *Yersinia pestis,* Lon protease upregulates secretion of type III effectors by degrading YmoA, a transcriptional repressor of the TTSS locus[72]. However, In *S. typhimurium*, Lon protease represses secretion by targeting the transcriptional activators HilC and HilD for degradation[73]. Though Lon depletion led to massive induction of apoptosis in macrophages and enhanced invasion of epithelial cells in vitro, mutants were severely attenuated in mouse models of infection, presumably due to increased recognition by immune effectors due to over-secretion of virulence factors.

Clp and HtrA proteases have also been implicated in the production and secretion of virulence factors in Gram positive organisms. Many syndromes caused by *S. aureus* result from a set of secreted toxins. One important protein, alpha-hemolysin, is a pore-forming toxin whose expression is dependent on the presence of ClpXP protease. Depletion of either clpX or clpP in *S. aureus* abolished transcription of the *hla* gene that encodes alpha-hemolysin, presumably by

stabilizing an as yet unidentified repressor, and abolished virulence[74]. Similarly, production of listeriolysin O, a major virulence factor of *L. monocytogenes*, is abolished in a clpP mutant, thus abrogating virulence in this organism[75].

In some bacteria, HtrA may itself be a secreted virulence factor. The HtrA protease from *Chlamydia trachomatis* is specifically secreted into the host cell cytosol during intracellular infection, though its host substrates are unknown. In the extracellular gut pathogen *Helicobacter pylori*, HtrA is a secreted virulence factor that cleaves host E-cadherin leading to disruptions of tight junctions in the gut epithelia, promoting the development of *H. pylori*-mediated peptic ulcer disease[25].

DEGRADATIVE PROTEASES AS TARGETS FOR THERAPEUTIC DEVELOPMENT

As outlined above, the degradative proteases are clearly required for many cellular processes. Despite their importance, conservation, and druggability, there are currently no commercial antibiotics that target this enzyme class in bacteria. In some cases, disruption of proteases should block cell growth; in others, modulation could limit host virulence. Targeting virulence pathways with small molecules represents a relatively new strategy in the development of antibiotics. The inhibition of virulence does not kill the organism during infection but prevents colonization of the host. Unable to establish a foothold in the host, bacteria targeted with anti-virulence antibiotics would be removed by the host immune system and other clearance mechanisms. The inhibition of Clp protease in the urinary tract pathogen, *Staphylococcus epidermidis*, is an ideal example of the potential efficacy of anti-virulence antibiotics. In *S. epidermidis,* Clp protease is not essential for normal growth, but does degrade the negative regulator Spx, a transcription factor that represses biofilm formation and the production of genes that facilitate primary attachment to the uroepithelium[66]. A small molecule that inhibited Clp protease function could prevent the ability the organism to colonize the urinary tract, and lead to clearance of the organism by the flow of urine.

The major benefit for anti-virulence strategies is thought to be the relatively low selective pressure that inhibiting virulence would exhibit at a population level[76]. Traditional antibiotics, which focused on targeting central processes within an organism, placed an enormous selective pressure on the population, which has partly contributed to the current pandemic of drug resistance. In theory, the inhibition of virulence would apply a milder evolutionary pressure, as most virulence pathways are not essential for normal growth. As no anti-virulence antibiotics currently exist on the market, however, it is not yet clear if these antibiotics will be less prone to the development of resistant strains in a clinical setting.

Additionally, protease inhibitors may prove valuable in combination therapy regimens, as antibiotics tend to cause stress and may make these degradative complexes even more important for survival. In mycobacteria, partial Clp depletion enhances killing by aminoglycosides, which increase the intracellular concentration of abnormal polypeptides and may amplify the demand for Clp protease to remove these aberrant products[4].

One of the most attractive properties of proteases as antibiotic targets is their druggability. The active sites of protease are among the most well-studied in enzymology. Numerous chemical moieties, such as the peptide boronates, aldehydes, β-lactones and epoxyketones, have been employed in the development of inhibitors of proteolytic activity[77]. Structurally analogous to actual protein substrates, these compounds either mimic the transition state of peptide hydrolysis or act as suicide inhibitors, forming stable complexes with the nucleophilic residue of the protease active site and preventing hydrolysis of normal substrates. Specificity of these compounds usually derives from the attached peptidyl backbone or adjacent peptidic structures. While the specific active site residues involved in peptide bond hydrolysis are similar among the different classes of proteases, amino acids in close proximity to the catalytic center, which form the protease specificity pocket, are much less conserved. These preferences can be identified via substrate

scanning using a large library of small peptides to determine the residues that best fit into pockets adjacent to the active site. As the development of the human proteasome inhibitor bortezomib illustrates, the peptides that are most efficiently recognized and cleaved can be linked to inhibitory groups to generate a specific active site inhibitor[78].

In the case of these large proteolytic complexes, there are also numerous other structural features beyond the protease active site that could serve as targets for modulators. The ATP-dependent proteases have Walker box ATP-binding domains that could be inhibited with nucleoside analogs. In theory, cofactor analogs could also be used to modulate the function of Lon protease, which has been shown to bind polyphosphate and DNA in order to facilitate targeted degradation of numerous substrates. And because every protease exists as a multimer and often must dock with separate ATPases or adapters to facilitate protein degradation, compounds that inhibit these interactions could also affect protease function.

While proteases can make very good targets, there are substantial problems that must be overcome. First and foremost is the issue of specificity. While the well-conserved active sites of the proteases make them extremely druggable targets, this same property presents a significant challenge to the development of specific modulators. As a result, promiscuous inhibitors that target broad classes of proteases have been described far more frequently than specific agents[79]. Second, the presence of homologous proteases in eukaryotes may result in significant toxicity from a protease modulator. Each of the degradative proteases in bacteria exists in the mitochondria of humans. Moreover, the human proteasome resembles the proteasome found in *M. tuberculosis*. In general, however, empirical evidence described below and the fact that a broad array of currently used antibiotics target bacterial enzymes that have homologous counterparts in humans suggests there to be enough divergence between bacterial and eukaryotic structures to enable the development of specific modulators.

Third, target-based approaches to antibacterial development have not been very successful, on the whole. The failure rate of rational drug design strategies for antibacterials when compared to chronic diseases or even viral drug development has largely been attributed to the complex issue of accessibility[80]. Antibiotics must first be able to penetrate the site of infection. Once there, a small molecule has to cross into the target bacterium, traversing various layers including the thick peptidoglycan layers of Gram positive organisms or the double membrane structure with intermingling peptidoglycan in Gram negative pathogens. Once at its target, the compound must avoid being inactivated by bacterial metabolic processes and exported by efflux pumps. It has not been easy to formulate a set of structural guidelines that fulfills all of these criteria, and the lack of rules has prevented targeted exploration of the ideal chemical space. Comparing the molecular weight and hydrophobicity of antibiotics to other commercially available pharmaceuticals reveals that that antimicrobials tend to have a higher mass and lower partition coefficients (defined as the ratio of concentrations of a compound in two phases of a mixture of octanol and water) then other developed drugs[81]. This may serve as evidence of the bias that the additional permeability obstacles have placed on attaining a successful development candidate for antibacterial drug design.

While traditional target-based approaches have hit roadblocks, our understanding of the intracellular proteolytic complexes could allow development of unique target-based whole cell screening methods. In the easiest iteration, endogenous substrates degraded by a particular protease could be fused to a fluorescent protein resulting in a reporter protein that could measure specific protease activity modulation in a whole cell system **(Figure 1.3, top)**. If done in duplex with a general whole cell screen, a library of compounds could be simultaneously tested for a general effect on bacterial growth and for activity through specific enzyme inhibition. In more complex forms, inhibition of adapter-protease interactions or the post-translational modification and subsequent degradation of critically regulated proteins could be tested by engineering a

reporter protein that was known to be modified through a specific mechanism **(Figure 1.3, bottom)**.

While this method has yet to be utilized in a productive manner, traditional whole cell and target-based approaches have already yielded potential lead compounds that modulate activity of the degradative proteases. The exploration of these enzymes as therapeutic targets remains in its infancy, but promising examples exist for Clp, Lon, the mycobacterial proteasome, and HtrA. **(Table 1.2)**.

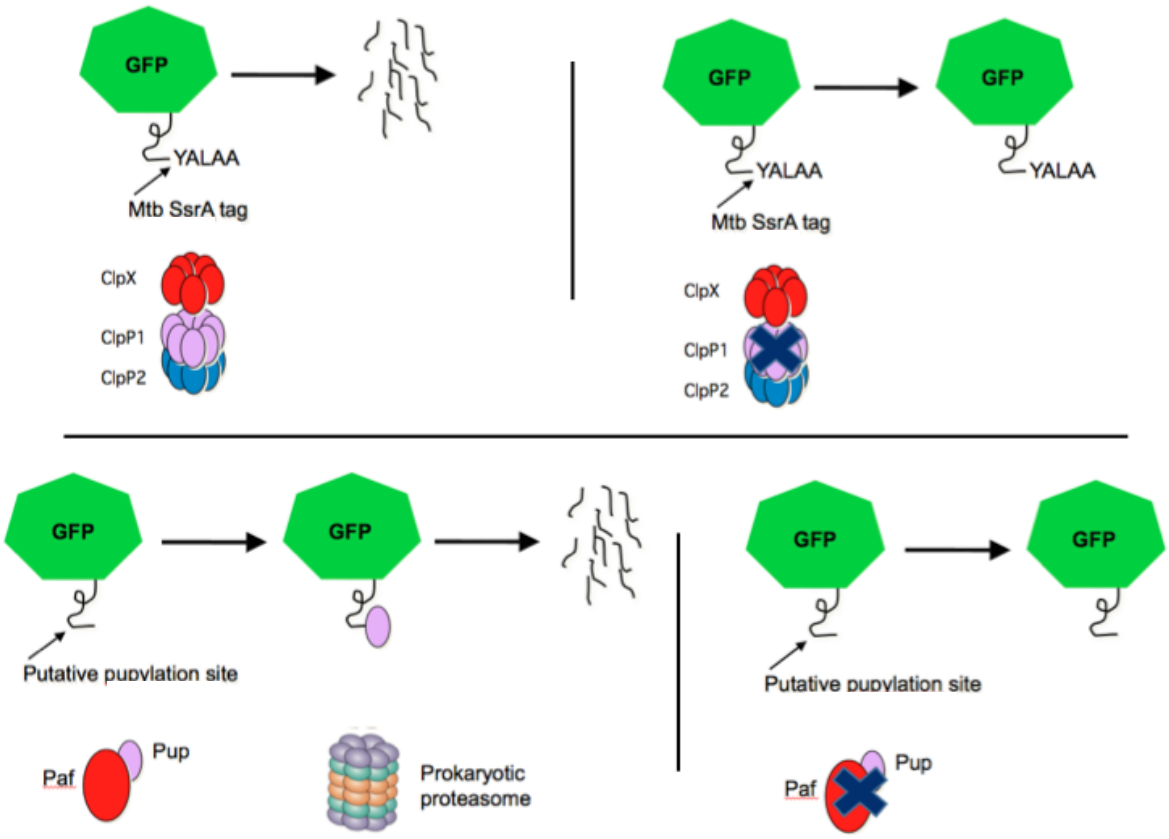
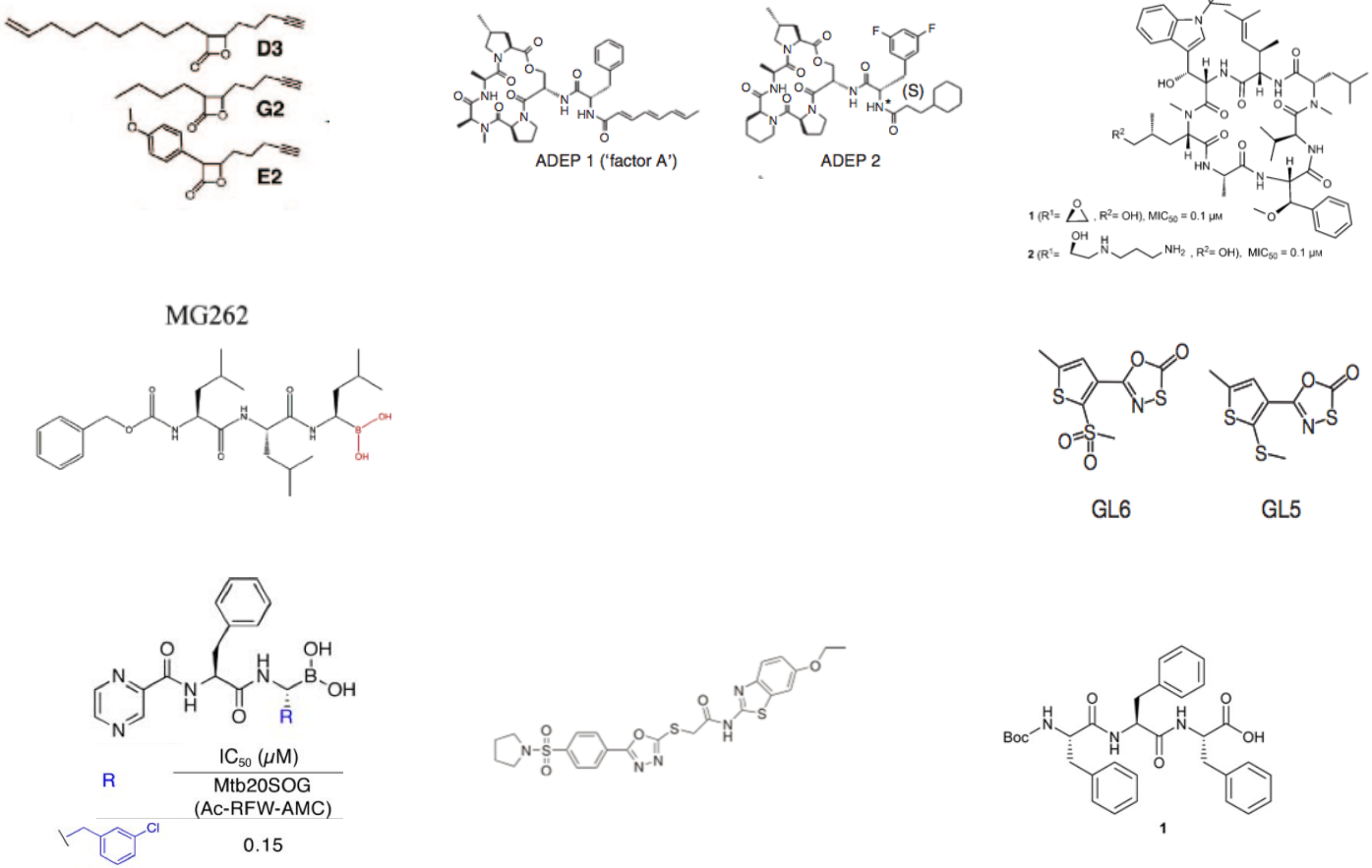


Figure 1.3 Targeted whole cell screens can identify modulators of proteolysis

Our understanding of substrate recognition and proteolysis by the degradative proteases can be harnessed to develop targeted whole cell screens that simultaneously identify cell-permeable small molecules that act in a certain pathway. For example, one could screen for inhibitors of Clp protease by fusing the SsrA tag to GFP. In wildtype cells expressing the fusion protein, fluorescence would be minimal (top left panel), but in the presence of Clp protease inhibitors, the GFP-SsrA would be stabilized and fluorescence would be preserved (top right panel). In a more complex screen, accessory pathways could also be exploited for drug development. In *M. tuberculosis*, the proteasome accessory factor (Paf) is responsible for the addition of the essential prokaryotic ubiquitin protein (Pup) to proteins destined for degradation by the prokaryotic proteasome. Though none have been reported to date, fusing a putative pupylation site to GFP would result in a strain that lacked fluorescence in an unperturbed state (bottom left panel), but would express stabilized GFP in presence of proteasome inhibitors or modulators of upstream proteins such as Paf or Pup (bottom right panel).

Table 1.2 – Chemical structures identified as bacterial-specific protease modulators



Small molecules that have been identified as protease modulators include ClpP inhibitors (beta-lactones, top left), ClpP activators (ADEPs, top middle), ClpC1 activators (cyclomarin, top right), Lon inhibitors (MG262, middle left), Mtb proteasome inhibitors (oxathiazol-2-ones GL6/ GL5, middle right; bortezomib derivatives, bottom left), HtrA inhibitors (HHI, bottom middle), and DegS activators (tripeptide compounds, bottom right).

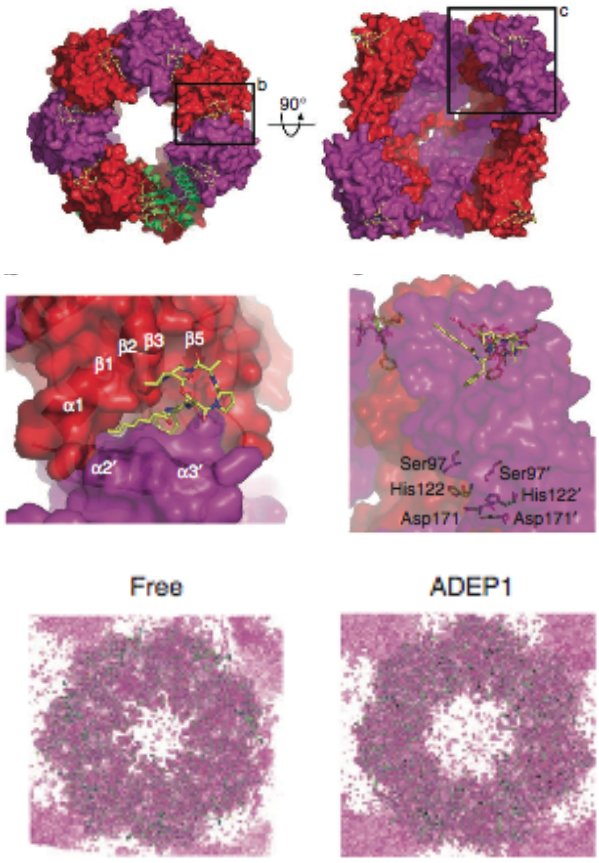
Clp protease

To date, most efforts to modulate the degradative proteases have concentrated on Clp protease, and have produced three classes of compounds with markedly different mechanisms of action. The first class, β-lactones, were initially identified as naturally occurring compounds, some of which exhibited anti-microbial activity[82]. A large class of bioactive compounds, the β-lactones appear to interact with a variety of enzymes. However, activity-based protein profiling (ABPP) using chemical probes derived from aromatic and aliphatic derivitized trans β-lactones demonstrated that these particular compounds were Clp-specific inhibitors in the pathogenic organisms *S. aureus* and *L. monocytogenes*[83,84]. As Clp is not essential in these organisms, the trans-β-lactones fail to inhibit growth, but do inhibit virulence factor production. While the efficacy of the trans β-lactones as anti-virulence compounds has been well-established, it would be interesting to determine their utility as traditional growth inhibitors in *M. tuberculosis*, the only pathogenic organism where Clp protease is required for growth in vitro and in vivo.

The second class of Clp-specific agents, ADEPs, were initially isolated from the fermentation broth of *Streptomyces hawaiiensis*, and found to have an inhibitory effect on growth in staphylococci, streptococci, and other Gram-positive organisms[85]. Sequencing of resistant mutants revealed that the potential target was Clp protease, but as Clp protease is not required for normal growth in these organisms, it was peculiar that ADEPs exhibited their bactericidal effect through an interaction with a dispensable protease[86]. Surprisingly, an in vitro peptidase assay using ClpP protease from *Bacillus subtilis* revealed that ADEPs activated proteolysis. ADEPs additionally prevented interaction between ClpP and cognate ATPase adapters, ClpX and ClpC, thus turning ClpP into an activated, unregulated proteolytic machine[87]. Though ADEP-activated ClpP compelxes were unable to degrade fully formed proteins, in vitro and in vivo transcription/translation assays demonstrated that the dysregulated enzyme cleaved newly forming polypeptides and unstructured proteins. Crystallization of ADEP-bound ClpP and X-ray analysis showed that ADEPs bound to ClpP at the same site where the IGF loops of ATPase

adapters interact when they activate the ClpP protease complex[88]. By binding at this site, which is formed by the interaction of two adjacent ClpP monomers in the assembled complex, ADEPs stabilize multimerization of ClpP subunits without the ATPases and reorient the N-terminal segment of ClpP **(Figure 1.4)**. Binding with ADEPs disrupt the N-terminal hydrophobic interactions that normally convert the pore of ClpP from a closed to an open form, capable of degrading unfolded proteins and peptides in an unregulated manner. Promising targets for further development, the in vivo efficacy of ADEPs has been demonstrated in mouse infections with *E. faecalis, S. pneumoniae,* and *S. aureus*.

Recently, the natural product cyclomarin was found to have bactericidal activity for *M. tuberculosis*. Sepharose-bound cyclomarin has been shown to bind to the essential ATPase adapter ClpC1[89]. Its mode of action is not completely clear, though early experiments suggest that it increases protein degradation *in vivo*. Thus, multiple components of the Clp protease could potentially serve as drug targets.



From Lee et al, 2010 [88]

Figure 1.4 Crystal structure of ADEP-*Bacillus subtilis* (Bs) ClpP interaction reveals ADEP-dependent activation mechanism
(Top left) The tetradecameric BsClpP-ADEP1 complex viewed along a sevenfold molecular symmetry axis. Monomers are alternatively colored red and magenta, with one monomer shown as green ribbon for clarity. Bound ADEP1 molecules are shown as stick models. (Top right) Side view of the BsClpP-ADEP1 interaction. (Middle left) Close up view of one of the 14 ADEP1 binding sites boxed in top left. The binding site is a complementary pocket composed by two adjacent subunits of ClpP. (Middle right) Close up view of boxed region in top right, with transparent molecular surface showing the catalytic triads and ADEP1 on the peripheral surface. (Bottom) Electron density map derived form the structures of free BsClpP1 (left) and BsClpP-ADEP1 (right) illustrating the axial pore of the ClpP proteolytic core.

Lon protease

Lon protease has been targeted using a mechanism-based design strategy. Taking advantage of the different cleavage specificity of the *S. typhimurium* and human mitochondrial homolog, a boronate-based compound was designed that specifically inhibits the bacterial enzyme[90]. In addition, a fluorescent dansyl moiety was conjugated to the compound to allow functional and enzymatic studies of inhibition. The resultant compound inhibited Lon protease purified from *S. typhimurium* via a two-step mechanism that required ATP binding. Initial binding of the boronate to the active site of Lon protease was followed by a conformational shift in the overall structure of the enzyme, leading to the stabilization of the transition state intermediate[91]. While this compound has yet to be developed further or tested in a whole bacterium or in vivo setting, it represents an important step forward in validating Lon protease as a drug target, especially in the case of *S. typhimurium,* where Lon protease regulates Type III secretion and is essential for full virulence. As overexpression of Lon protease is lethal in *E. coli*, activators may also provide tractable therapeutic targets, though none have yet been described.

Mycobacterial Proteasome

The proteasome of *M. tuberculosis* is an attractive target for drug development, given that the mycobacterial proteasome is essential for both virulence and survival under nitric oxide stress. Initial efforts to develop proteasome-targeting antibiotics exploited the differential substrate specificity between the eukaryotic and bacterial proteins. A single amino acid change in the eukaryotic boronate-based proteasome inhibitor, bortezomib, resulted in a compound with eight-fold selectivity for the *M. tuberculosis* enzyme[92]. In order to obtain more specific and selective agents, an in vitro screen of 20,000 compounds identified two oxathiazol-2-one inhibitors that were highly specific for the *M. tuberculosis* proteasome, and had no measurable inhibitory activity on the eukaryotic homolog[93]. Co-crystallization has shown that the specificity of these compounds lies in their interactions with non-conserved residues far outside of the active site that are brought close to the oxathiazol-2-one after compound-induced conformational changes.

Moreover, the compounds bind covalently to the active site, acting as irreversible inhibitors. These compounds are bactericidal against *M. tuberculosis* in culture, both in replicating and non-replicating conditions. As non-dividing *M. tuberculosis* may play an important role in the ability of bacteria to survive in the host for decades in a latent state, agents such as these could be very attractive weapons to add to the antituberculous arsenal.

HtrA protease

The HtrA proteases are the most accessible of the degradative proteases as they lie outside of the bacterial cytoplasm. Using structure-based virtual screening, a set of compounds were discovered that had the potential to act as small molecule inhibitors of *Helicobacteri pylori* HtrA, the enzyme that cleaves host E-cadherins[94]. Of the 22 compounds retrieved from virtual screening, six were shown to inhibit cleavage of E-cadherins by Hp HtrA in an *in vitro* proteolysis assay. One compound, HHI, was shown to preserve E-cadherin adhesions in an epithelial monolayer model of *H. pylori* infection. HHI is extremely specific for *H. pylori* HtrA and does not inhibit the human HtrA1 or a panel of other host proteases[25].

In a manner analogous to ADEPs, activation of HtrA protease might also provide an avenue to effective antibiotics. In fact, tripeptidic small molecules are able to activate DegS protease of *E. coli*. Tripeptides trigger DegS activation by binding to the PDZ domain via hydrophobic interactions, a manner similar to potential substrates[95]. Though we do not yet know if these molecules have activity against whole cells, targeting outside of the active site is a promising strategy to develop new, specific modulators.

CONCLUDING REMARKS

The degradative proteases offer a set of attractive targets for the development of antibacterial drugs. In most cases, they fill roles that are required for growth and survival, either *in vitro* or

during infection. Even “nonessential” proteases can be converted to toxic enzymes by activators such as ADEPs. If problems with specificity and accessibility can be overcome, the medicinal chemistry experience with protease inhibitors could be harnessed to exploit a plethora of vulnerable targets.

In fact, understanding the role and mechanism of these proteases has already revealed additional potential targets beyond the active sites of these enzymes. These include residues that are exposed upon conformational changes in the enzyme (e.g., the oxathiazol-2-one inhibitors of the bacterial proteasome), active sites in accessory enzymes (e.g., cylcomarin binding the ClpC1 ATPase), regulatory regions of proteins (e.g, peptidic inhibitors that activate DegS through its PDZ domain), and interacting surfaces between members of a proteolytic complex (e.g., ADEPs in preventing interactions between ClpP and the ATPases). Certainly, there may be more. For example, essential proteases are essential because certain substrates must be degraded. Agents that specifically blocked degradation of these substrates, perhaps through interaction with protein-specific degradation adapters, should have similar activity to inhibitors of the proteases themselves. Identifying candidates for such approaches will require a deeper understanding of the physiological roles of many of the degradative proteases, and elucidation of each proteases “degradome,” or the entire set of proteins and pathways regulated by controlled proteolysis. Furthering our understanding of substrate degradation by these proteases can facilitate the development of innovative whole-cell target based screens for the identification of small molecule modulators.

Despite the need for further studies to systematically understand the functional significance of the degradative proteases, we already have many important tools to discover and design new compounds to modulate their function. Partial crystal structures, which exist for every enzyme, have already proven invaluable for designing inhibitors and accelerating the optimization of drug candidates by medicinal chemistry. *In vitro* assays of these enzymes can easily be turned into

screens for new inhibitors, and fusion of *in vivo* substrates to fluorescent proteins could even permit the design of whole-cell, pathway-based assays for compounds that not only modulate activity but also are capable of accessing the proteases in intact bacteria.

Considerable challenges remain. Inhibitors of mitochondrial function, such as the many antibiotics that block protein synthesis at prokaryotic ribosomes, are often toxic, particularly upon prolonged exposure. In many ways, the most significant obstacle may be designing specific inhibitors that don't inhibit human homologs, although a precedent has been set with the development of the oxathiazol-2-one inhibitors of the mycobacterial proteasome. Additionally, many protease modulators are peptidic small molecules, which have poor pharmacologic properties, often lacking oral bioavailability with poor entry through the bacterial cell wall and degradation by peptidases in the host and bacterial cytoplasm. However, the combination of a deep understanding of the biochemistry and physiology of these proteins with the expertise that has come from developing a number of successful protease inhibitor drugs already makes the degradative proteases very compelling targets for antibacterial drug development.

WORKS CITED

1. Poordad, F. *et al.* Boceprevir for untreated chronic HCV genotype 1 infection. *N Engl J Med* **364**, 1195–1206 (2011).

2. Drag, M. & Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov* **9**, 690–701 (2010).

3. Plaut, A. G. The IgA1 proteases of pathogenic bacteria. *Annu Rev Microbiol* **37**, 603–622 (1983).

4. Raju, R. M. *et al.* Mycobacterium tuberculosis ClpP1 and ClpP2 Function Together in Protein Degradation and Are Required for Viability in vitro and During Infection. *PLoS Pathog* **8**, e1002511 (2012).

5. Schmidt, R., Bukau, B. & Mogk, A. Principles of general and regulatory proteolysis by AAA+ proteases in Escherichia coli. *Research in Microbiologoy* **160**, 629–636 (2009).

6. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**, 27–43 (1999).

7. Burton, B. M. & Baker, T. A. Remodeling protein complexes: insights from the AAA+ unfoldase ClpX and Mu transposase. *Protein Sci* **14**, 1945–1954 (2005).

8. Gur, E. & Sauer, R. T. Recognition of misfolded proteins by Lon, a AAA+ protease. *Genes Dev* **22**, 2267–2277 (2008).

9. Iniesta, A. A., McGrath, P. T., Reisenauer, A., McAdams, H. H. & Shapiro, L. A phospho-signaling pathway controls the localization and activity of a protease complex critical for bacterial cell cycle progression. *Proc Natl Acad Sci USA* **103**, 10935–10940 (2006).

10. Schlothauer, T., Mogk, A., Dougan, D. A., Bukau, B. & Turgay, K. MecA, an adaptor protein necessary for ClpC chaperone activity. *Proc Natl Acad Sci USA* **100**, 2306–2311 (2003).

11. Kress, W., Maglica, Z. & Weber-Ban, E. Clp chaperone-proteases: structure and function. *Res Microbiol* **160**, 618–628 (2009).

12. Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. The N-end rule in bacteria. *Science* **254**, 1374–1377 (1991).

13. Erbse, A. *et al.* ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature* **439**, 753–756 (2006).

14. Herman, C., Thévenet, D., Bouloc, P., Walker, G. C. & D'Ari, R. Degradation of carboxy-terminal-tagged cytoplasmic proteins by the Escherichia coli protease HflB (FtsH). *Genes Dev* **12**, 1348–1355 (1998).

15. Akiyama, Y., Kihara, A., Tokuda, H. & Ito, K. FtsH (HflB) is an ATP-dependent protease selectively acting on SecY and some other membrane proteins. *J Biol Chem* **271**, 31196–31201 (1996).

16. Kihara, A., Akiyama, Y. & Ito, K. A protease complex in the Escherichia coli plasma membrane: HflKC (HflA) forms a complex with FtsH (HflB), regulating its proteolytic activity against SecY. *EMBO J* **15**, 6122–6131 (1996).

17. Rotanova, T. V. *et al.* Slicing a protease: Structural features of the ATP-dependent Lon proteases gleaned from investigations of isolated domains. *Protein Sci* **15**, 1815–1828 (2006).

18. Nomura, K., Kato, J., Takiguchi, N., Ohtake, H. & Kuroda, A. Effects of inorganic polyphosphate on the proteolytic and DNA-binding activities of Lon in Escherichia coli. *J Biol Chem* **279**, 34406–34410 (2004).

19. Yoo, S. J. *et al.* Mutagenesis of two N-terminal Thr and five Ser residues in HslV, the proteolytic component of the ATP-dependent HslVU protease. *FEBS Letters* **412**, 57–60 (1997).

20. Bochtler, M. *et al.* The structures of HsIU and the ATP-dependent protease HsIU-HsIV. *Nature* **403**, 800–805 (2000).

21. Dahlmann, B. *et al.* The multicatalytic proteinase (prosome) is ubiquitous from eukaryotes to archaebacteria. *FEBS Letters* **251**, 125–131 (1989).

22. Darwin, K. H., Lin, G., Chen, Z., Li, H. & Nathan, C. F. Characterization of a Mycobacterium tuberculosis proteasomal ATPase homologue. *Molecular Microbiology* **55**, 561–571 (2004).

23. Pearce, M. J., Mintseris, J., Ferreyra, J., Gygi, S. P. & Darwin, K. H. Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis. *Science* **322**, 1104–1107 (2008).

24. Spiess, C., Beil, A. & Ehrmann, M. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97**, 339–347 (1999).

25. Hoy, B. *et al.* Helicobacter pylori HtrA is a new secreted virulence factor that cleaves E-cadherin to disrupt intercellular adhesion. *EMBO reports* **11**, 798–804 (2010).

26. Kim, D. Y. & Kim, K. K. Structure and function of HtrA family proteins, the key players in protein quality control. *J. Biochem. Mol. Biol.* **38**, 266–274 (2005).

27. Krojer, T. *et al.* Structural basis for the regulated protease and chaperone function of DegP. *Nature* **453**, 885–890 (2008).

28. Carroll, P., Faray-Kele, M.-C. & Parish, T. Identifying vulnerable pathways in Mycobacterium tuberculosis using a knock-down approach. *Applied and Environmental Microbiology* 1–19 (2011).doi:10.1128/AEM.02880-10

29. Tomoyasu, T. *et al.* The Escherichia coli FtsH protein is a prokaryotic member of a protein family of putative ATPases involved in membrane functions, cell cycle control, and gene expression. *J Bacteriol* **175**, 1344–1351 (1993).

30. Wilken, C., Kitzing, K., Kurzbauer, R., Ehrmann, M. & Clausen, T. Crystal structure of the DegS stress sensor: How a PDZ domain recognizes misfolded protein and activates a protease. *Cell* **117**, 483–494 (2004).

31. Akiyama, Y., Kihara, A. & Ito, K. Subunit a of proton ATPase F0 sector is a substrate of the FtsH protease in Escherichia coli. *FEBS Letters* **399**, 26–28 (1996).

32. Withey, J. H. & Friedman, D. I. A salvage pathway for protein structures: tmRNA and trans-translation. *Annu Rev Microbiol* **57**, 101–123 (2003).

33. Shi, W. *et al.* Pyrazinamide inhibits trans-translation in Mycobacterium tuberculosis. *Science* **333**, 1630–1632 (2011).

34. Lindquist, S. & Craig, E. A. The heat-shock proteins. *Annu Rev Genet* **22**, 631–677 (1988).

35. Isaac, D. D., Pinkner, J. S., Hultgren, S. J. & Silhavy, T. J. The extracytoplasmic adaptor protein CpxP is degraded with substrate by DegP. *Proc Natl Acad Sci USA* **102**, 17775–17779 (2005).

36. Raivio, T. L. & Silhavy, T. J. Transduction of envelope stress in Escherichia coli by the Cpx two-component system. *J Bacteriol* **179**, 7724–7733 (1997).

37. Grigorova, I. L. *et al.* Fine-tuning of the Escherichia coli sigmaE envelope stress response relies on multiple mechanisms to inhibit signal-independent proteolysis of the

transmembrane anti-sigma factor, RseA. *Genes Dev* **18**, 2686–2697 (2004).

38. Bury-Moné, S. *et al.* Global Analysis of Extracytoplasmic Stress Signaling in Escherichia coli. *PLoS Genet* **5**, e1000651 (2009).

39. Humphreys, S. *et al.* Role of the Two-Component Regulator CpxAR in the Virulence of Salmonella enterica Serotype Typhimurium. *Infection and Immunity* **72**, 4654–4661 (2004).

40. Manganelli, R., Voskuil, M. I., Schoolnik, G. K. & Smith, I. The Mycobacterium tuberculosis ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Molecular Microbiology* **41**, 423–437 (2001).

41. Vijayakumar, S. R. V., Kirchhof, M. G., Patten, C. L. & Schellhorn, H. E. RpoS-regulated genes of Escherichia coli identified by random lacZ fusion mutagenesis. *J Bacteriol* **186**, 8499–8507 (2004).

42. Zhou, Y. & Gottesman, S. Regulation of proteolysis of the stationary-phase sigma factor RpoS. *J Bacteriol* **180**, 1154–1158 (1998).

43. Becker, G., Klauck, E. & Hengge-Aronis, R. Regulation of RpoS proteolysis in Escherichia coli: the response regulator RssB is a recognition factor that interacts with the turnover element in RpoS. *Proc Natl Acad Sci USA* **96**, 6439–6444 (1999).

44. Bougdour, A., Cunning, C., Baptiste, P. J., Elliott, T. & Gottesman, S. Multiple pathways for regulation of σS (RpoS) stability in Escherichia coli via the action of multiple anti-adaptors. *Molecular Microbiology* **68**, 298–313 (2008).

45. Wang, X. *et al.* Antitoxin MqsA helps mediate the bacterial general stress response. *Nat Chem Biol* **7**, 359–366 (2011).

46. Kim, Y. *et al.* Escherichia coli toxin/antitoxin pair MqsR/MqsA regulate toxin CspD. *Environmental Microbiology* **12**, 1105–1121 (2010).

47. Langklotz, S. & Narberhaus, F. The Escherichia coli replication inhibitor CspD is subject to growth-regulated degradation by the Lon protease. *Molecular Microbiology* **80**, 1313–1325 (2011).

48. Gerth, U. *et al.* Clp-dependent proteolysis down-regulates central metabolic pathways in glucose-starved Bacillus subtilis. *J Bacteriol* **190**, 321–331 (2008).

49. Kuroda, A. *et al.* Role of inorganic polyphosphate in promoting ribosomal protein degradation by the Lon protease in E. coli. *Science* **293**, 705–708 (2001).

50. Stephani, K., Weichart, D. & Hengge, R. Dynamic control of Dps protein levels by ClpXP and ClpAP proteases in Escherichia coli. *Molecular Microbiology* **49**, 1605–1614 (2003).

51. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. & Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**, 671–683 (2003).

52. Tatsuta, T. *et al.* Heat shock regulation in the ftsH null mutant of Escherichia coli: dissection of stability and activity control mechanisms of sigma32 in vivo. *Molecular Microbiology* **30**, 583–593 (1998).

53. Griffith, K. L., Shah, I. M. & E Wolf, R., Jr Proteolytic degradation of Escherichia coli transcription activators SoxS and MarA as the mechanism for reversing the induction of the superoxide (SoxRS) and multiple antibiotic resistance (Mar) regulons. *Molecular Microbiology* **51**, 1801–1816 (2004).

54. Festa, R. A. *et al.* Prokayrotic ubiquitin-like protein (Pup) proteome of Mycobacterium tuberculosis. *PLoS ONE* **5**, e8589 (2010).

55. Pruteanu, M. & Baker, T. A. Controlled degradation by ClpXP protease tunes the levels of the excision repair protein UvrA to the extent of DNA damage. *Molecular Microbiology* **71**, 912–924 (2009).

56. Frank, E. G., Ennis, D. G., Gonzalez, M., Levine, A. S. & Woodgate, R. Regulation of SOS mutagenesis by proteolysis. *Proc Natl Acad Sci USA* **93**, 10291–10296 (1996).

57. Quon, K. C., Yang, B., Domian, I. J., Shapiro, L. & Marczynski, G. T. Negative control of bacterial DNA replication by a cell cycle regulatory protein that binds at the chromosome origin. *Proc Natl Acad Sci USA* **95**, 120–125 (1998).

58. Iniesta, A. A. & Shapiro, L. A bacterial control circuit integrates polar localization and proteolysis of key regulatory proteins with a phospho-signaling cascade. *Proc Natl Acad Sci USA* **105**, 16602–16607 (2008).

59. Camberg, J. L., Hoskins, J. R. & Wickner, S. ClpXP protease degrades the cytoskeletal protein, FtsZ, and modulates FtsZ polymer dynamics. *Proc Natl Acad Sci USA* **106**, 10614–10619 (2009).

60. Dziedzic, R. *et al.* Mycobacterium tuberculosis ClpX interacts with FtsZ and interferes with FtsZ assembly. *PLoS ONE* **5**, e11058 (2010).

61. Wu, W. F., Zhou, Y. & Gottesman, S. Redundant in vivo proteolytic activities of Escherichia coli Lon and the ClpYQ (HslUV) protease. *J Bacteriol* **181**, 3681–3687 (1999).

62. Ohlsen, K. L., Grimsley, J. K. & Hoch, J. A. Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase. *Proc Natl Acad Sci USA* **91**, 1756–1760 (1994).

63. Le, A. T. T. & Schumann, W. The Spo0E phosphatase of Bacillus subtilis is a substrate of the FtsH metalloprotease. *Microbiology (Reading, Engl)* **155**, 1122–1132 (2009).

64. Cutting, S. *et al.* SpoVM, a small protein essential to development in Bacillus subtilis, interacts with the ATP-dependent protease FtsH. *J Bacteriol* **179**, 5534–5542 (1997).

65. Liu, J., Cosby, W. M. & Zuber, P. Role of lon and ClpX in the post-translational regulation of a sigma subunit of RNA polymerase required for cellular differentiation in Bacillus subtilis. *Molecular Microbiology* **33**, 415–428 (1999).

66. Wang, C. *et al.* Role of spx in biofilm formation of Staphylococcus epidermidis. *FEMS Immunol Med Microbiol* **59**, 152–160 (2010).

67. Brondsted, L., Andersen, M. T., Parker, M., Jorgensen, K. & Ingmer, H. The HtrA Protease of Campylobacter jejuni Is Required for Heat and Oxygen Tolerance and for Optimal

Interaction with Human Epithelial Cells. *Applied and Environmental Microbiology* **71**, 3205–3212 (2005).

68. Lewis, C. *et al.* Salmonella enterica Serovar Typhimurium HtrA: regulation of expression and role of the chaperone and protease activities during infection. *Microbiology* **155**, 873–881 (2009).

69. Alix, E. & Blanc-Potard, A.-B. MgtC: a key player in intramacrophage survival. *Trends Microbiol* **15**, 252–256 (2007).

70. Alix, E. & Blanc-Potard, A.-B. Peptide-assisted degradation of the Salmonella MgtC virulence factor. *EMBO J* **27**, 546–557 (2008).

71. Gandotra, S., Lebron, M.B. & Ehrt, S. The *Mycobacterium tuberculosis* proteasome active site threonine is essential for persistence yet dispensable for replication and resistance to ntiric oxide. *PLoS Pathog* **6**, e1001040 (2010).

72. Jackson, M. W., Silva-Herzog, E. & Plano, G. V. The ATP-dependent ClpXP and Lon proteases regulate expression of the Yersinia pestis type III secretion system via regulated proteolysis of YmoA, a small histone-like protein. *Molecular Microbiology* **54**, 1364–1378 (2004).

73. Takaya, A., Kubota, Y., Isogai, E. & Yamamoto, T. Degradation of the HilC and HilD regulator proteins by ATP-dependent Lon protease leads to downregulation of Salmonella pathogenicity island 1 gene expression. *Molecular Microbiology* **55**, 839–852 (2004).

74. Frees, D., Qazi, S. N. A., Hill, P. J. & Ingmer, H. Alternative roles of ClpX and ClpP in Staphylococcus aureus stress tolerance and virulence. *Molecular Microbiology* **48**, 1565–1578 (2003).

75. Gaillot, O., Pellegrini, E., Bregenholt, S., Nair, S. & Berche, P. The ClpP serine protease is essential for the intracellular parasitism and virulence of Listeria monocytogenes. *Molecular Microbiology* **35**, 1286–1294 (2000).

76. Rasko, D. A. & Sperandio, V. Anti-virulence strategies to combat bacteria-mediated disease. *Nat Rev Drug Discov* **9**, 117–128 (2010).

77. de Bettignies, G. & Coux, O. Proteasome inhibitors: Dozens of molecules and still counting. *Biochimie* **92**, 1530–1545 (2010).

78. Goldberg, A. Bortezomib's Scientific Origins and Its Tortuous Path to the Clinic. *Bortezomib in the Treatment of Multiple Myeloma* (2011).

79. Supuran, C. T., Scozzafava, A. & Clare, B. W. Bacterial protease inhibitors. *Med Res Rev* **22**, 329–372 (2002).

80. Coates, A. R. M. & Hu, Y. Novel approaches to developing new antibiotics for bacterial infections. *British Journal of Pharmacology* **152**, 1147–1154 (2007).

81. Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29–40 (2007).

82. Böttcher, T. & Sieber, S. A. Beta-lactones as privileged structures for the active-site labeling of versatile bacterial enzyme classes. *Angewandte Chemie (International ed in English)* **47**, 4600–4603 (2008).

83. Böttcher, T. & Sieber, S. A. Beta-lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of Staphylococcus aureus. *J Am Chem Soc* **130**, 14400–14401 (2008).

84. Böttcher, T. & Sieber, S. A. Beta-lactones decrease the intracellular virulence of Listeria monocytogenes in macrophages. *ChemMedChem* **4**, 1260–1263 (2009).

85. Brötz-Oesterhelt, H. *et al.* Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nat Med* **11**, 1082–1087 (2005).

86. Gominet, M., Seghezzi, N. & Mazodier, P. Acyl depsipeptide (ADEP) resistance in Streptomyces. *Microbiology* (2011).doi:10.1099/mic.0.048454-0

87. Kirstein, J. *et al.* The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. *EMBO Mol Med* **1**, 37–49 (2009).

88. Lee, B.-G. *et al.* Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nat Struct Mol Biol* **17**, 471–478 (2010).

89. Schmitt, E. K. *et al.* The Natural Product Cyclomarin Kills Mycobacterium Tuberculosis by Targeting the ClpC1 Subunit of the Caseinolytic Protease. *Angewandte Chemie (International ed in English)* (2011).doi:10.1002/anie.201101740

90. Frase, H., Hudak, J. & Lee, I. Identification of the proteasome inhibitor MG262 as a potent ATP-dependent inhibitor of the Salmonella enterica serovar Typhimurium Lon protease. *Biochemistry* **45**, 8264–8274 (2006).

91. Frase, H. & Lee, I. Peptidyl boronates inhibit Salmonella enterica serovar Typhimurium Lon protease by a competitive ATP-dependent mechanism. *Biochemistry* **46**, 6647–6657 (2007).

92. Lin, G., Tsu, C., Dick, L., Zhou, X. K. & Nathan, C. Distinct specificities of Mycobacterium tuberculosis and mammalian proteasomes for N-acetyl tripeptide substrates. *J Biol Chem* **283**, 34423–34431 (2008).

93. Lin, G. *et al.* Inhibitors selective for mycobacterial versus human proteasomes. *Nature* **461**, 621–626 (2009).

94. Löwer, M. *et al.* Inhibitors of Helicobacter pylori protease HtrA found by “virtual ligand” screening combat bacterial invasion of epithelia. *PLoS ONE* **6**, e17986 (2011).

95. Hauske, P. *et al.* Peptidic small molecule activators of the stress sensor DegS. *Mol. BioSyst.* **5**, 980–985 (2009).

Chapter 2:

The requirements of Clp protease in *Mycobacterium tuberculosis* during normal growth and infection

This chapter has been published as a research article:

Raju RM*, Unnikrishnan M*, Rubin DHF, Krishnamoorthy V, Kandror O, Akopian TN, Goldberg AL, Rubin EJ. *Mycobacterium tuberculosis* ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLoS Pathogens* **8**, e1002511 (2012).

* Authors contributed equally to this work.

Section 2.1 – Chapter Overview and Attributions

In this chapter, I present our initial efforts to determine the requirements of Clp protease in mycobacteria. As mentioned above, the study of the Clp protease complex in mycobacteria is warranted for several reasons. First, Clp is dispensable for normal growth in model organisms where it has been extensively studied, but a genome wide screen for essential genes conducted in our laboratory suggested that the protease was absolutely required for normal growth in Mtb. Second, most organisms possess one *clpP* gene, which forms the proteolytic core of the enzyme, but mycobacteria encode two homologous genes, *clpP1* and *clpP2*, situated in a single operon. Through a combination of genetic and biochemical studies, I present below, we were able to show that both *clpP1* and *clpP2* are independently essential for normal growth in *Mycobacterium smegmatis* (Msm) and that the Mtb forms of ClpP1 and ClpP2 interact together to form a single proteolytic core, denoted ClpP1P2 **(APPENDIX 1)**. Taking advantage of this, we expressed active site mutants of ClpP1 in Mtb, which had a slight dominant negative effect on growth in vitro. These mutants, however, were highly attenuated in a mouse model of infection, suggesting that Clp protease has a role in infection as well. Scratching the surface of a functional understanding of Clp, we also show that Clp-dependent recognition of SsrA tagged proteins is conserved in mycobacteria, and that Clp plays a role in the degradation of abnormal protein products that accumulate in the cytoplasm. Together, these findings demonstrate that the Clp protease plays a critical role in mycobacterial physiology.

Attributions. The manuscript in this chapter was published with both Meera Unnikrishnan and myself as co- authors. Meera pioneered the Clp project in our laboratory and laid the initial groundwork for the studies presented here. She constructed the pTtet_*clpP2* conditional mutant described in Figure 2.2D and 2.2E, and conducted the experiments in Figure 2.3A. I constructed all other strains used in this chapter, and performed all experiments independently with the exception of Figure 2.1A-D, which was done with tremendous help from Tatos Akopian and Olga

Kandror in the laboratory of Alfred Goldberg. I wrote the manuscript with editorial input from all authors on the paper.

Section 2.2 – *Mycobacterium tuberculosis* ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection

Ravikiran M. Raju[a,2], Meera Unnikrishnan [a,2], Daniel H.F. Rubin [a], Vidhya Krishnamoorthy [a], Olga Kandror [b], Tatos N. Akopian [b], Alfred L. Goldberg [b], Eric J. Rubin [a,1]

[a] Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02138 USA

[b] Department of Cell Biology, Harvard Medical School, Boston, MA 02138 USA

[1] Corresponding author: erubin@hsph.harvard.edu, ph: (617) 432-3337

[2] These authors contributed equally to this work.

ABSTRACT

In most bacteria, Clp protease is a conserved, non-essential serine protease that regulates the response to various stresses. Mycobacteria, including *Mycobacterium tuberculosis* (Mtb) and *Mycobacterium smegmatis*, unlike most well studied prokaryotes, encode two ClpP homologs, ClpP1 and ClpP2, in a single operon. Here we demonstrate that the two proteins form a mixed complex (ClpP1P2) in mycobacteria. Using two different approaches, promoter replacement, and a novel system of inducible protein degradation, leading to inducible expression of *clpP1* and *clpP2*, we demonstrate that both genes are essential for growth and that a marked depletion of either one results in rapid bacterial death. ClpP1P2 protease appears important in degrading missense and prematurely terminated peptides, as partial depletion of ClpP2 reduced growth specifically in the presence of antibiotics that increase errors in translation. We further show that the ClpP1P2 protease is required for the degradation of proteins tagged with the SsrA motif, a tag

co-translationally added to incomplete protein products. Using active site mutants of ClpP1 and ClpP2, we show that the activity of each subunit is required for proteolysis, for normal growth of Mtb *in vitro* and during infection of mice. These observations suggest that the Clp protease plays an unusual and essential role in Mtb and may serve as an ideal target for antimycobacterial therapy.

AUTHOR SUMMARY

Due to the significant and rapid rise in multidrug resistant *Mycobacterium tuberculosis* (Mtb), there is an urgent need to validate novel drug targets for the treatment of tuberculosis. Here, we show that Clp protease is an ideal potential target. Mtb encodes two ClpP genes, ClpP1 and ClpP2, which associate together to form a single proteolytic complex, referred to as ClpP1P2. Both proteins are required for growth *in* vitro and in a mouse model of infection. Depletion of either protein results in rapid death of the bacteria. Interestingly, this is rare among bacteria, most of which have only one ClpP gene that is dispensable for normal growth. We also show that Clp protease plays an important quality control role by clearing abnormally produced proteins. As known antimycobacterial therapeutics increase errors in protein synthesis, inhibitors of ClpP1P2 protease in Mtb may prove synergistic with already existing agents.

INTRODUCTION

Intracellular protein degradation is critical for maintaining cellular homeostasis through protein quality control and regulation of numerous biological pathways[1,2]. In eukaryotes, the ubiquitin-proteasome pathway constitutes the predominant degradation pathway[3]. Most prokaryotes, however, possess a variety of ATP-dependent serine protease complexes, such as Lon and Clp protease[4], and some actinomycetes and archaea contain proteasomes, which are threonine proteases. Interestingly, *Mycobacterium tuberculosis* (Mtb) encodes both a proteasome and Clp protease. While recent work has explored the role of the Mtb proteasome[5-7], little is known about

mycobacterial Clp protease. This serine protease was first discovered and is best characterized in *Escherichia coli*[8,9]. The Clp proteolytic complex is formed by the association of proteolytic subunits, ClpP, with ATPase adapters, ClpX or ClpA in Gram-negative organisms and ClpX or ClpC in Gram-positive organisms. *E. coli* ClpP is a tetradecamer composed of two stacked heptameric rings of identical ClpP subunits that form an internal proteolytic chamber[10]. This core associates with distinct hexameric ATPase adapters, ClpX and ClpC1 in mycobacteria, which provide substrate specificity and catalyze ATP-dependent unfolding of globular proteins[11,12]. In *E. coli*, the ClpXP protease is involved in the regulation of the DNA damage response and degradation of SsrA-tagged peptides stalled on the ribosome[13,14]. Clp proteolytic enzymes are also required for full virulence in several pathogenic organisms, including *Listeria monocytogenes* where the protease is required for the production of α-listeriolysin[15,16]. In most bacteria including *E. coli*, Clp protease is dispensable for normal growth, and in fact, until recently, the only organism in which *clpP* has been found to be essential is *Caulobacter crescentus*, where Clp degrades CtrA, an inhibitor of cell cycle progression[17].

Unlike most bacteria, which have a single ClpP subunit, the genome of Mtb encodes two closely related ClpP homologs, *clpP1* and *clpP2*, in a single operon. A transposon-based mutagenesis screen for essential genes in Mtb predicted that ClpP2 and the ATPase adapters ClpC1 and ClpX, were required for normal growth[18] while a recent publication has shown that ClpP1 is essential[19]. Here, we show that both ClpP1 and ClpP2 are required for growth, and that their activity is important for the removal of abnormal proteins. Our data suggest that ClpP1 and ClpP2 assemble to form a single proteolytic complex, referred to as ClpP1P2, that is required for normal growth *in vitro* and during infection. In related studies, we have found that although pure ClpP1 and ClpP2 by themselves form tetradecamers, they are inactive. However, in the presence of low molecular weight activators they reassociate to form a mixed tetradecamer, ClpP1P2, which is capable of proteolysis[20]. The unusual properties of this heteromeric complex, the absence of such

an enzyme in the eukaryotic cytoplasm, and the essentiality of both subunits make ClpP1P2 protease an attractive target for novel therapeutic development for the treatment of tuberculosis.

RESULTS

ClpP1 and ClpP2 subunits interact to form a single proteolytic complex

Mycobacterial genomes contain two homologous ClpP protease genes, *clpP1* and *clpP2*, arranged in a putative operon. To investigate whether the two proteins may function together in a complex, we co-expressed Mtb *clpP1* and *clpP2*, each containing a different C terminal epitope tag, in *Mycobacterium smegmatis* (Msm). We used affinity chromatography with nickel resin to isolate 6x-His tagged Mtb ClpP2 together with associated proteins from the Msm cell lysate. As shown in **Figure 2.1A**, a fraction of the c-myc tagged ClpP1 bound to the Ni column and co-eluted with ClpP2. To verify that ClpP1 and ClpP2 co-eluted from the Ni column may be associated in a complex, we applied the fraction from the Ni column containing both proteins to an anti-c-myc agarose column and analyzed by SDS PAGE. **Figure 2.1B** shows that a large fraction of the ClpP2 was associated with ClpP1. Incidentally, expression of the Mtb proteins in Msm also led to the co-isolation of Msm ClpP1 and ClpP2, as shown by tandem mass spectrometry of the purified complex. In each case, peptides present uniquely in Mtb or Msm ClpP1 and ClpP2 were detected **(Figure 2.1C)**.

If ClpP1 and ClpP2 do in fact associate to form a single proteolytic core, we reasoned that mutations blocking the catalytic activity of one subunit might reduce that activity of the enzyme. We identified likely active site residues of ClpP1 and ClpP2 by mapping the Mtb proteins onto *E. coli* ClpP and locating the catalytic triad of Asp-His-Ser, which is characteristic of serine proteases. In both cases, the serine likely to be responsible for nucleophillic attack was replaced by an alanine (ClpP1 S98A and ClpP2 S110A). To analyze the effects of these mutations, we expressed and purified 6xHis-tagged forms of each protein, and assayed their effect on the enzymatic activity of the wild type ClpP1P2 in an *in vitro* peptidase assay[20]. Enzyme activity of the

reconstituted ClpP1P2 complex was quantified using cleavage of the fluorescence reporter, Z-Gly-Gly-Leu-AMC. As seen in **Figure 2.1D**, addition of an excess of mutated ClpP1 or ClpP2 to the active wild type ClpP1P2 complex inhibited proteolytic cleavage of a fluorescent peptide substrate, presumably by replacing the wild type subunits. These results suggest that the ClpP1 and ClpP2 subunits interact to form a single proteolytic complex *in vitro*, that each active site is important for activity, and that these mutations can be used as dominant negative inhibitors.

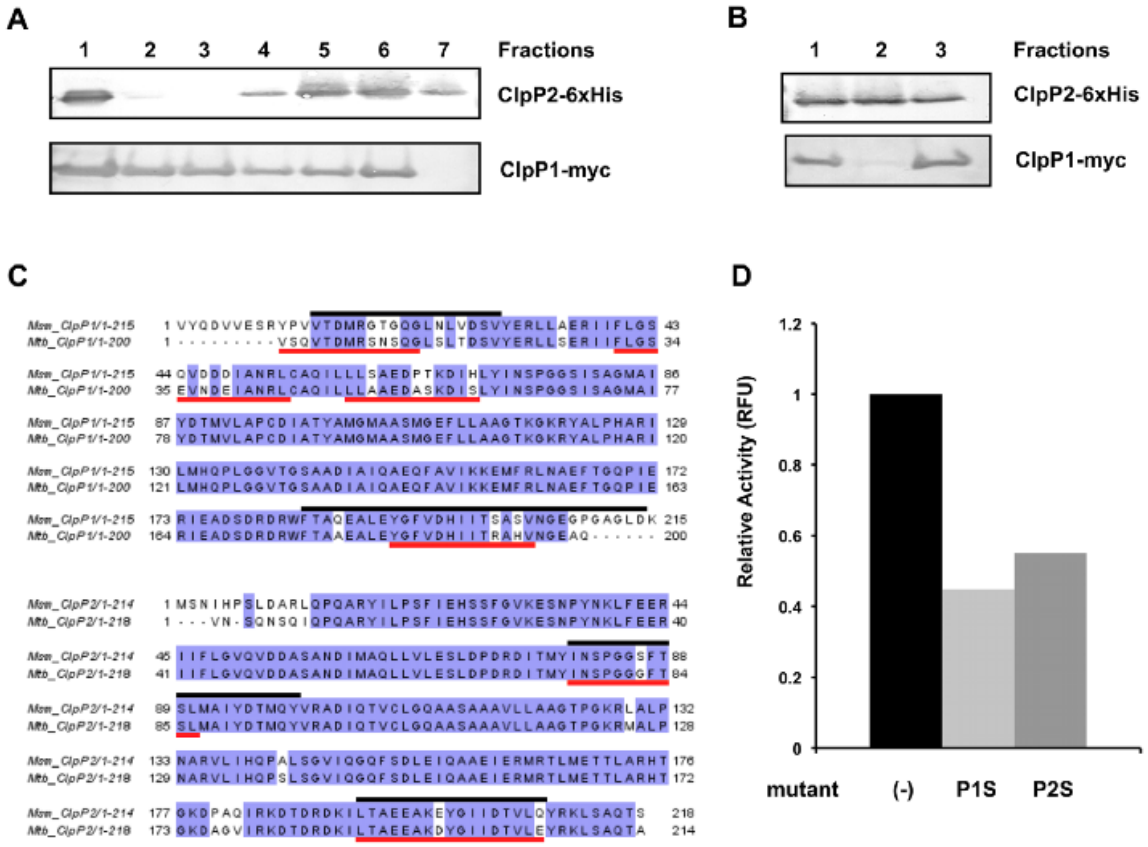


Figure 2.1 Mtb ClpP1 and ClpP2 interact, forming a multi-component protease, and share substantial similarity with ClpP1 and ClpP2 homologs in Msm.

(A) C-terminally myc-tagged Mtb ClpP1 and 6xHis-tagged Mtb ClpP2 were expressed in Msm. Lysate (lane 1) was prepared and loaded onto a Ni-column. After washing with PBS (lanes 2,3), Ni-bound material was eluted with 50 mM (lane 4), 100 mM (lane 5), 250 mM (lane 6, 7) of imidazole in PBS, and analyzed by immunoblotting using anti α-myc and α-6xHis antibodies.

(B) Fraction 6 from (A) was applied to an anti-myc column (lane 1). The flow through (lane 2), and bound material (lane 3) were analyzed by immunoblot with α-myc and α-His antibodies. Bound material was released from the anti-myc agarose beads by boiling in Laemmli buffer after washing with PBS.

(C) Bands representing ClpP1 and ClpP2 from (B) were sequenced by MS/MS revealing the presence of both Mtb and Msm homologs. Msm specific peptides are indicated by black lines, those specific to Mtb are indicated by red lines.

(D) Cleavage of fluorescent peptide Z-Gly-Gly-Leu-AMC was measured in the presence of 1 mg ClpP1, 1 mg Clp2, and the activating peptide Z-Leu-Leu (see accompanying paper). Addition of 5 mg of catalytically inactive mutants of either ClpP1 (ClpP1S) or ClpP2 (ClpP2S) markedly inhibited cleavage by the ClpP1P2 protease. Results graphed are a representative sample of results obtained.

Both ClpP1 and ClpP2 are required for normal growth *in vitro*

We employed three complementary strategies to determine if ClpP1 and ClpP2 are required for normal growth in mycobacteria. First, using mycobacterial recombineering[21], we replaced the endogenous promoter of *clpP1* and *clpP2* in Msm with a tetracycline-inducible promoter **(Figure 2.2A)**. Introduction of a tetracycline repressor resulted in a strain (ptet_clpP1P2) that could only be maintained in the presence of the inducer anhydrotetracycline (ATc) **(Figure 2.2B)**. In the absence of this compound, growth did not occur, but could be restored by the presence of an episomal plasmid containing both *clpP1* and *clpP2*. Plasmids expressing only *clpP1* or c*lpP2* alone could not rescue growth and depletion of either subunit resulted in bacterial death **(Figure 2.2C)**. Since complementation was conducted with Mtb homologs and subunits from different species associate into a functional tetradecamer, the ClpP1P2 complex is likely very similar in Msm and Mtb. Furthermore, active site mutants of either ClpP1 or ClpP2 were unable to complement ptet_clpP1P2 in the absence of ATc, suggesting that the activity of both subunits were required for normal growth (data not shown).

Second, we inserted a tetracycline inducible promoter upstream of the *clpP1P2* operon via homologous recombination in Msm creating a strain in which *clpP2* was inducibly expressed **(Figure 2.2D),** and *clpP1* was under the control of its native promoter (ptet_clpP2). In accord with the previous findings, the growth of this strain was dramatically inhibited in the absence of ATc **(Figure 2.2E)**.

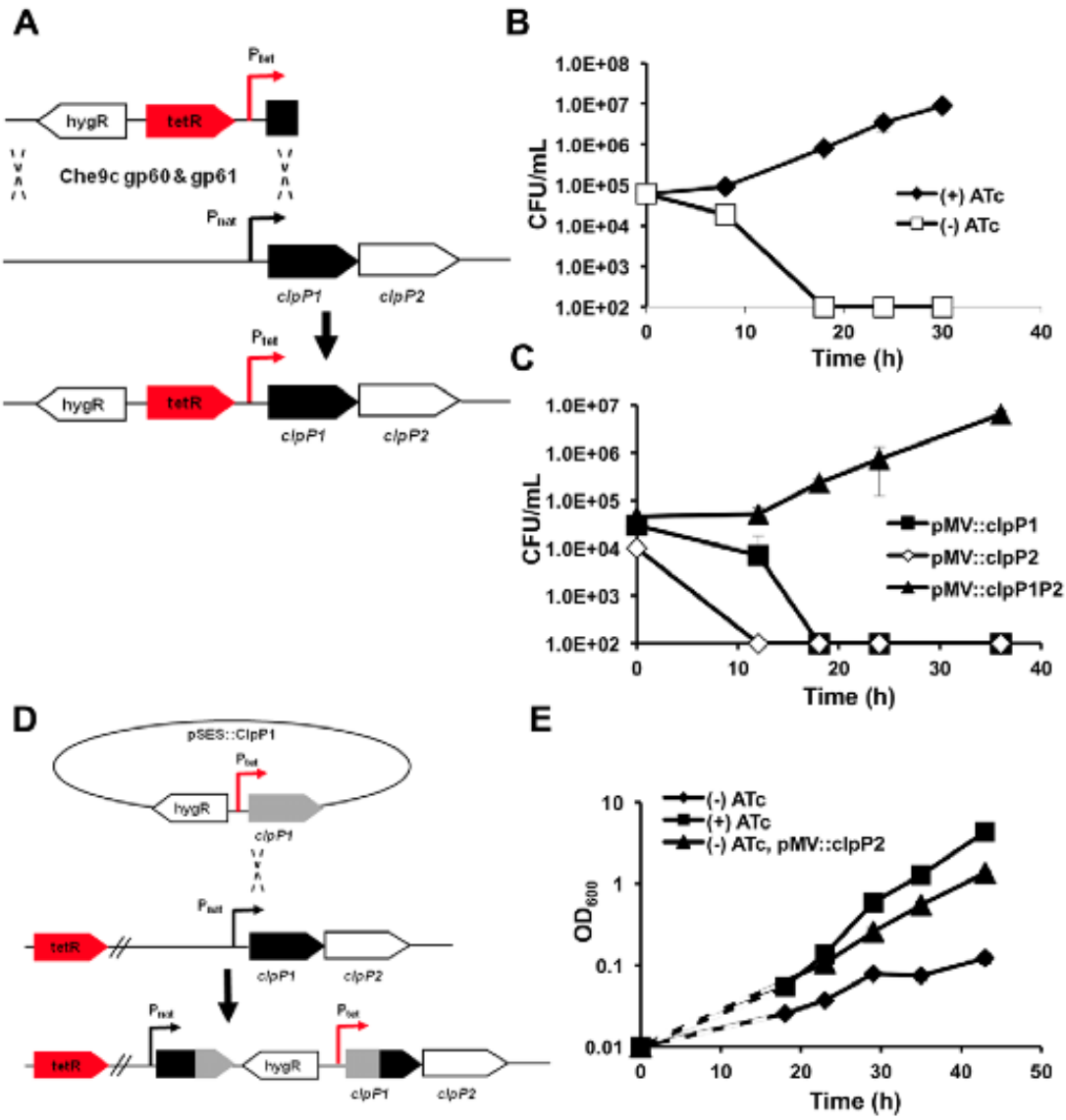


Figure 2.2 Both ClpP1 and ClpP2 are essential for normal growth in mycobacteria

(A) Schematic representation of mycobacterial recombineering, employed to replace the endogenous promoter of the clpP1P2 operon with a ATc-inducible promoter (Msm strain ptet_clpP1P2).

(B) Growth curves of Msm ptet_clpP1P2 in the presence (50 ng/mL) or absence of inducer ATc. Data are represented as mean CFU/mL +/- standard deviation.

(C) Growth curves of Msm ptet_clpP1P2 complemented with clpP1, clpP2 or both clpP1 and clpP2 in the absence of inducer ATc. Data are represented as mean CFU/mL +/- standard deviation.

(D) Schematic representation of genetic strategy used to create a tetracycline inducible conditional Msm ClpP2 mutant (Msm strain ptet_ClpP2)

(E) Growth curves of Msm ptet_clpP2 in the presence (50 ng/mL) or absence of inducer ATc. Msm ptet_clpP2 was also complemented with clpP2 in the absence of ATc. Data are represented as mean OD_{600} +/- standard deviation. Dashed lines represent assumed growth rates until first measured growth point.

Third, we used a system of inducible protein degradation recently developed in Msm **(Figure 2.3A)**[22]. Briefly, we employed mycobacterial recombineering to add an inducible degradation (ID) tag to the C-terminus of ClpP2 (clpP2_ID). Upon cleavage of the tag by a tetracycline inducible HIV-2 protease, an SsrA sequence is revealed on the substrate that directs degradation of the protein. By inserting epitope tags C-terminally to the HIV-2 protease recognition motif (FLAG) and N-terminally to the SsrA tag (c-myc), we were able to monitor the amount of ClpP2 by immunoblotting. As shown in **Figure 2.3B**, induction of HIV-2 protease resulted in degradation of the majority of ClpP2 and inhibited bacterial growth **(Figure 2.3C)**. Using this system, we did not observe cell death, perhaps due to incomplete inhibition, as would be expected for a system where the protease targets itself. Loss of ClpP2, as measured by immunoblotting, was rapid and reached near completion within hours. Furthermore, the growth defect was complemented by expression of Mtb *clpP2* using a constitutively active promoter. A similar approach with ClpP1 was unsuccessful as extended C-terminal tagging was not tolerated, and the ID tag was indiscriminately cleaved. Collectively, these results confirm that both ClpP1 and ClpP2 are required for normal growth in mycobacteria, presumably because they function together in the ClpP1P2 complex.

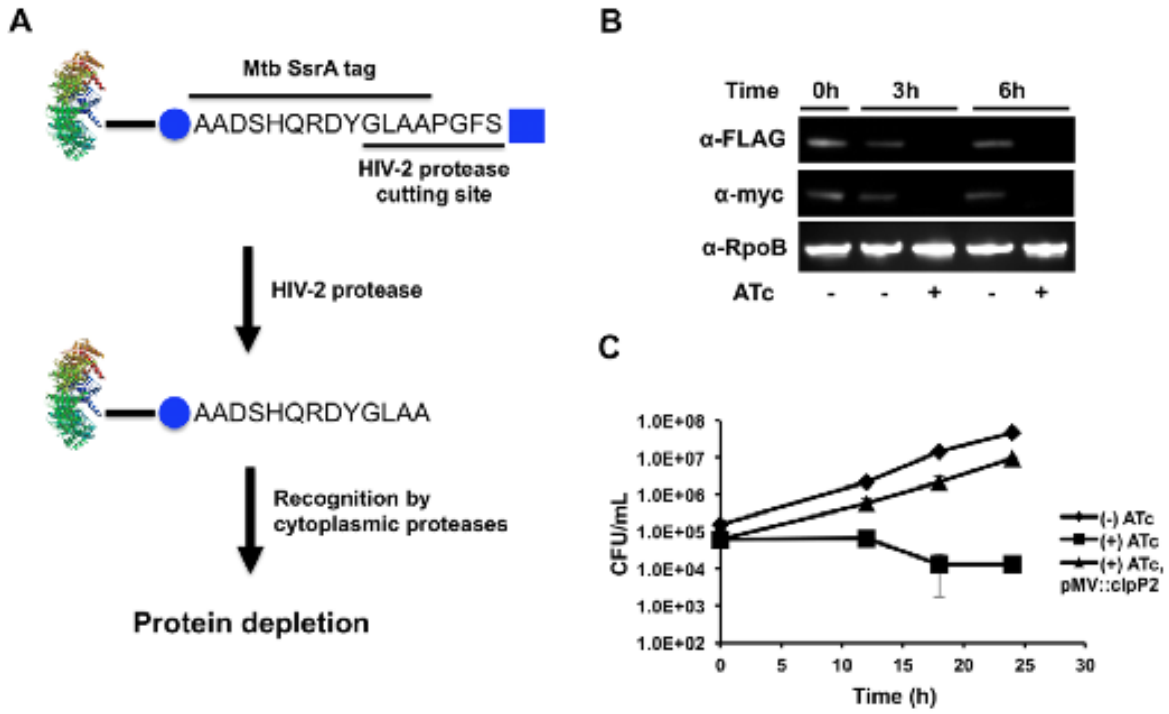


Figure 2.3 Inducible protein degradation demonstrates requirement of ClpP2 for normal growth

(A) Schematic representation of the inducible degradation system used to inducibly deplete ClpP2 (Msm strain clpP2_ID). Induction of HIV-2 protease with ATc leads to cleavage of the HIV-2 protease cutting site and exposure of a SsrA tag on the tagged protein. Cleavage by HIV-2 protease and subsequent degradation can be tracked via the FLAG (square) and c-myc (circle) epitope tags, respectively, included on the inducible degradation tag.

(B) Degradation of ClpP2 in clpP2_ID was tracked by Western in the absence or presence of inducer ATc. Blots were probed α-FLAG (loss indicates HIV-2 protease cleavage), α-myc (loss indicates target degradation), and α-RpoB (loading control).

(C) Growth curves of Msm clpP2_ID in the absence or presence (50 ng/mL) of inducer ATc. Msm clpP2_ID was also complemented with clpP2 in the presence of ATc. Data are represented as mean CFU/mL +/- standard deviation.

Mycobacterial Clp protease plays a role in protein quality control

In other bacteria, Clp plays a role in degrading abnormal proteins such as SsrA-tagged peptides that stall on the ribosome[23]. To determine the importance of ClpP1P2 protease in the degradation of misfolded proteins, we used antibiotics that alter protein synthesis in distinct ways including chloramphenicol, which blocks protein elongation without increasing mistranslation rates[24], and streptomycin and amikacin, which induce translational errors resulting in missesnse or prematurely-terminated polypeptides[25]. We found that the strain ptet_clpP2, in which *clpP2* expression is regulated by anhydrotetracycline, grows well in low or high concentrations of ATc, 1 to 100 ng/mL **(Figure 2.4A)**. Treatment with sublethal concentrations of chloramphenicol resulted in no difference in viability between bacteria maintained on low or high concentrations of ATc **(Figure 2.4A, bottom)**. In contrast, sub-MIC concentrations of the aminoglycosides streptomycin and amikacin significantly inhibited the growth of strains incubated in low concentrations of ATc, while they had no effect on growth of the strain maintained in high concentrations of ATc **(Figure 2.4A, top)**. Together, these results suggest that ClpP1P2 protease protects against error-prone translation by catalyzing the degradation of misfolded proteins.

To specifically assess whether ClpP1P2 is responsible for the removal of SsrA-tagged proteins in mycobacteria, we fused the mycobacterial SsrA-tag to the C-terminus of GFP (GFP-SsrA) and expressed the construct constitutively on an episomal plasmid. This construct was introduced into the strain clpP2_ID, in which ClpP2 degradation was regulated. In the presence of ClpP2 (and in wild type cells), there was no detectable GFP-SsrA. However, upon depletion of ClpP2, there was substantial accumulation of GFP-SsrA, as measured by both fluorescence and immunoblot analysis after four hours **(Figure 2.4B, 2.4C)**. Quantitative PCR showed that the rise of GFP-SsrA was not due to transcriptional activation of the gene **(Figure 2.4D)**. GFP lacking the SsrA tag is present at similar levels in all strains (data not shown). Because we cannot detect GFP-SsrA in the presence of Clp activity, we were unable to accurately measure changes in protein stability. However, the rate of accumulation of GFP-SsrA was consistent with the time course of ClpP2

depletion, which occurred over six hours, as shown by immunoblotting. Thus, functional ClpP1P2 protease is vital for the rapid clearance of SsrA-tagged substrates in mycobacteria.

Functional Clp protease is required for growth of Mtb *in vitro* and during infection

As shown above, catalytically inactive forms of ClpP1 and ClpP2 inhibit proteolysis by the wild type enzyme, possibly by replacement of wild type subunits with inactive ones. To assess whether ClpP1P2 activity is required for the growth of Mtb, we expressed a catalytically inactive form of Mtb *clpP1*, *clpP1 S98A*, on a tetracycline-inducible plasmid in wild type Mtb. Addition of ATc led to expression of the catalytically inactive mutant protein and resulted in a significant inhibition of growth **(Figure 2.5A)** while overexpression of wild type Mtb *clpP1* had no effect.

To determine if the dominant negative mutant of ClpP1 affected ClpP1P2 function during infection, we infected mice with a 3:1 mixture of Mtb expressing *clpP1 S98A* on a hygromycin-resistant doxycycline inducible plasmid and wild type Mtb (containing a kanamycin-resistant control vector). Mice were fed either normal chow or chow infused with the inducer doxycycline. Growth of Mtb was monitored by assessing CFU in lung tissue at day 30 post-infection. While there were no differences in the growth of wild type Mtb between treated and untreated mice, expression of the active site mutant significantly inhibited growth **(Figure 2.5B)**. Our results suggest that functional ClpP1P2 protease is required for the growth of Mtb both *in vitro* and during infection.

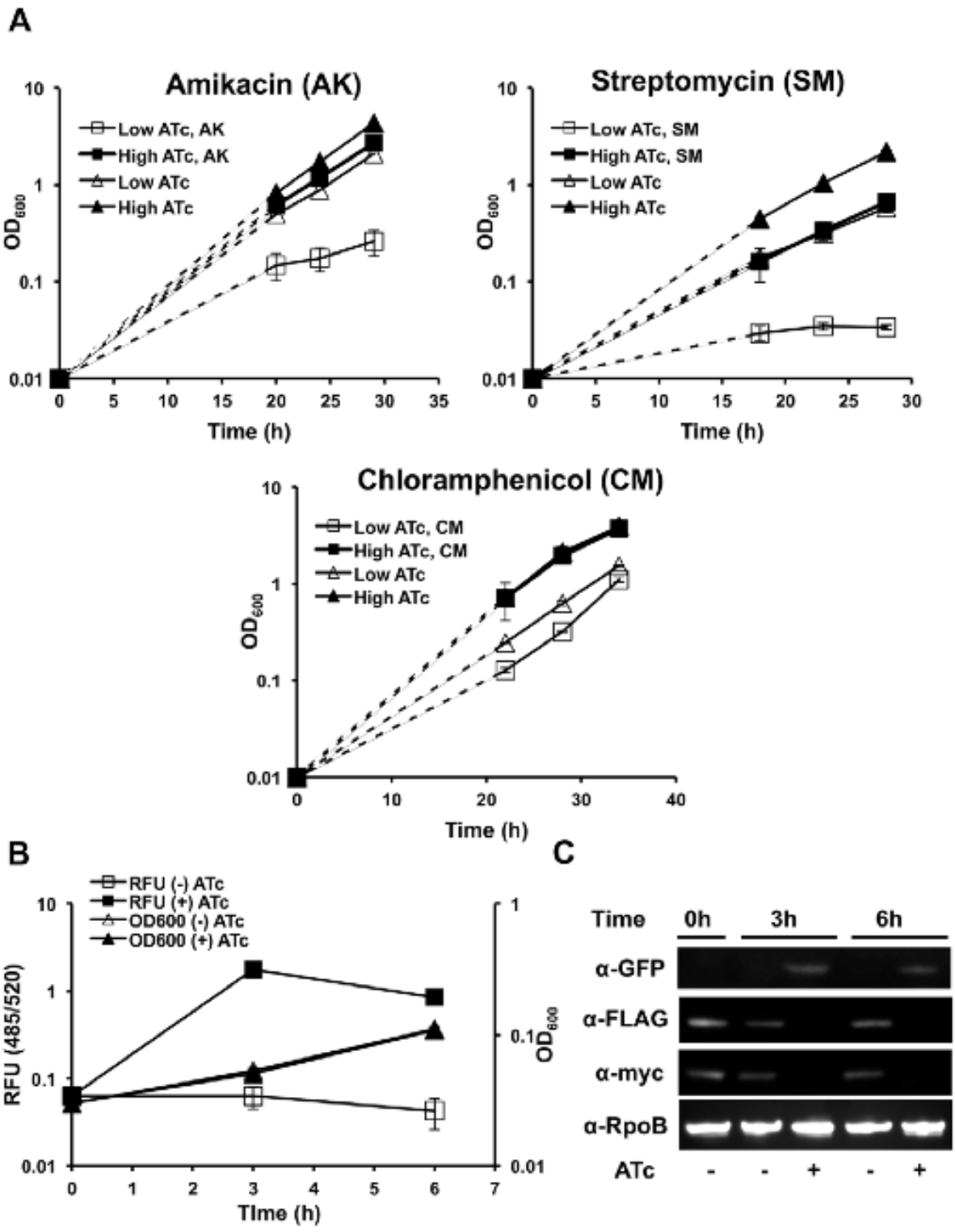


Figure 2.4 Clp protease is required for degradation of abnormal proteins and SsrA-tagged proteins in mycobacteria

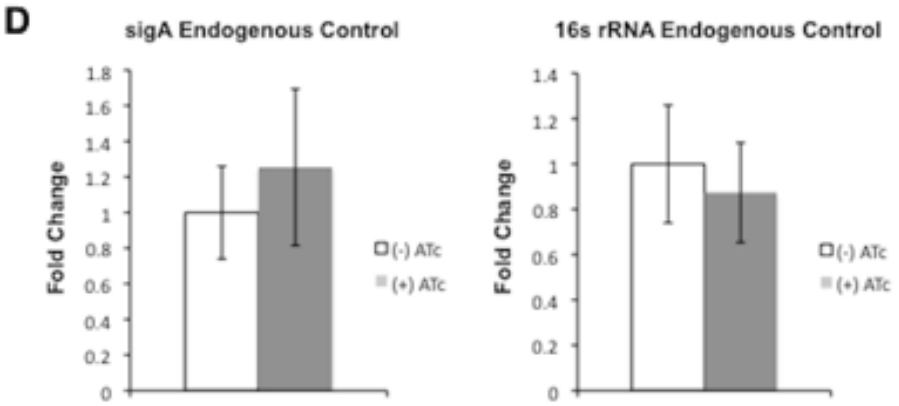


Figure 2.4 (continued) Clp protease is required for degradation of abnormal proteins and SsrA-tagged proteins in mycobacteria

(A) Growth curves of Msm ptet_clpP2 in growth medium containing low (1 ng/mL) or high (100 ng/mL) concentrations of inducer ATc, in the presence of either no drug, or amikacin (top left, 0.03 µg/mL), streptomycin (top right, 0.125 µg/mL), and chloramphenicol (bottom, 7.5 µg/mL). Data are represented as mean OD_{600} +/- standard deviation. Dashed lines represent assumed growth rates until first measured growth point.

(B) Increase in fluorescence (RFU, 485/520) and initial growth curve (OD_{600}) of Msm clpP2_ID expressing the fusion construct GFP-SsrA on a constitutively expressing plasmid, in the presence and absence of inducer, ATc. Data are represented as mean RFU or OD_{600} +/- standard deviation.

(C) Depletion of ClpP2 and increase in GFP-SsrA in Msm clpP2_ID expressing the fusion construct GFP-SsrA on a constitutively expressing plasmid was tracked by immunoblot. Blots were probed α-GFP, α-myc, α-FLAG, and α-RpoB (loading control).

D) Quantitative PCR of clpP2_ID was carried out to determine if increase in GFP-SsrA was due to transcriptional activation. RNA was isolated from clpP2_ID four hours after induction with ATc (+ ATc), and a culture of equal OD_{600} that was left uninduced (- ATc). Using both sigA (left) and 16s rRNA (right) as endogenous controls, there was no significant difference in transcription of GFP-SsrA between induced and uninduced cultures. Data are represented as mean fold change +/- standard deviation, with values normalized to those of the uninduced culture.

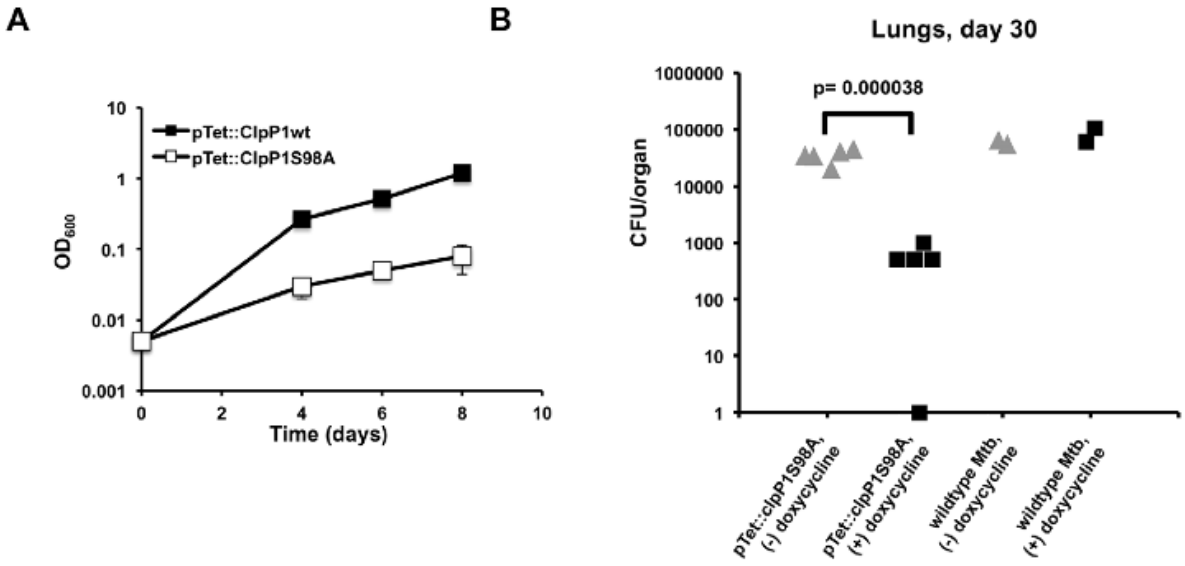


Figure 2.5 A catalytically inactive ClpP alleles inhibit Mtb growth *in vitro and during infection*

(A) Growth curves for Mtb overexpressing wild type ClpP1 or ClpP1 S98A via an ATc-inducible expression vector. Data are represented as mean OD_{600} +/- standard deviation. Dashed lines represent assumed growth rates until first measured growth point.

(B) Growth of Mtb containing a doxycycline-inducible plasmid expressing the mutant allele ClpP1 S98A in lungs of C57BL/6 mice 30 days post aerosol infection. Mice were infected via aerosol with a 3:1 mixture of mutant and wild type bacteria. Mice were fed either with chow containing (dark squares, N=5 mice) or lacking (gray triangles, N=5 mice) the inducer doxycycline. As a control, wild type Mtb was co-infected, and representative CFU/organs for the control are represented (right). Each point represents calculated total CFU/organ for each mouse. Not all mice received enough wild type bacteria to quantitate.

DISCUSSION

We find that the mycobacterial ClpP1P2 protease has two unusual properties that distinguishes it from other members of the prokaryotic ClpP family. First, the protease consists of distinct types of subunits, each of which is required for full activity of a single proteolytic complex. While other species do encode multiple ClpP subunits, two different proteolytic subunits forming a single protease has not been documented. Second, unlike in most bacteria that have been studied, ClpP1P2 activity is absolutely required for normal growth. This requirement is particularly striking as mycobacteria contain several cytoplasmic ATP-dependent proteolytic complexes, including FtsH, and the proteasome[7,26,27]. Clearly, the mycobacterial ClpP1P2 proteolytic core has unique roles that are important for viability.

The ClpP proteases that have been characterized biochemically in other bacteria and mitochondria are tetradecameric complexes containing a single type of proteolytic subunit. In mycobacteria, however, two different protein species contribute to protease activity. Although Mtb ClpP1 forms a tetradecameric complex, a crystal structure of Mtb ClpP1 lacks appropriate active site geometry to support proteolysis[28]. The presence of two ClpP subunits with distinct substrate preferences may facilitate an expansion of the peptide specificity of the complex, much like the eukaryotic proteasome. Interestingly, the Mtb proteasome is composed of a single type of subunit, and the presence of distinct subunits comprising a single proteolytic core is rare among prokaryotes.

There is at least one example of an essential role for ClpP. In *Caulobacter crescentus,* ClpXP degrades CtrA, a protein that normally inhibits cell cycle progression during cellular replication[29]. In this case, a single protein target is responsible for the essentiality of the enzyme. ClpP1P2 protease might play a similar role in mycobacteria. While this may be true, screens for essential proteins in mycobacteria suggest that, in addition to the clpP1 and clpP2, multiple Clp-associated ATPase adapters (clpX and clpC1) are also essential[18]. The requirement of multiple adapters

makes it possible that accumulation of multiple protein substrates contribute to the poor growth phenotype observed on depleting the ClpP1 and ClpP2 subunits in mycobacteria.

ClpP1P2 might be important for other reasons. As shown here, ClpP1P2 protease is required for the clearance of SsrA-tagged proteins. These tagged polypeptides are generated under conditions when protein synthesis is stalled and are required for ribosome release. In the absence of ClpP1P2-mediated proteolysis, protein synthesis might eventually be inhibited. In addition, ClpP1P2 protease is necessary for degrading abnormal proteins, such as those produced in the presence of certain antibiotics. Accumulation of such non-functional misfolded proteins might result in cellular stress in the absence of an effective system for their removal[30]. Clearance of damaged proteins might be particularly important in Mtb during infection, when cells are exposed to multiple oxidative and nitrosoative radicals that can induce protein damage. In fact, a transcriptional activator of the *clpP1P2* operon, clgR, is critically activated upon reaeration of hypoxic Mtb and during Mtb growth within the macrophage[31,32]. Degradation of pre-existing proteins during such stressful transitions may be the initial event that triggers adaptation and facilitates the bacterium's ability to handle a wide array of environmental challenges. Using a dominant negative overexpression mutant in Mtb, we have confirmed that optimal Clp proteolytic activity is required for growth during infection.

The essential nature of ClpP1P2 protease makes it an attractive target for antibiotic development, particularly because the proteases as a class are druggable enzymes and have already been validated as therapeutic targets in the treatment of HIV, hepatitis, and cancer[33]. In organisms where ClpP is not essential, uncontrolled activation of ClpP activity can be toxic. For example, in *E. coli*, acyldepsipeptide compounds reorganize the ClpP proteolytic core, causing dissociation from ATPase adapters, and indiscriminate protein degradation[34]. Compounds that produce a similar effect should result in toxicity in a broad range of organisms. In fact, it was recently discovered that the natural product cyclomarin kills Mtb by targeting the ClpC1 ATPase and

presumably increasing Clp-mediated proteolysis, as demonstrated in a whole cell fluorescence-based assay[35]. In mycobacteria, where ClpP1P2 protease activity is required and depletion of either subunit is bactericidal, either non-specific activation or inhibition could effectively limit bacterial growth. An example of a ClpP inhibitor with potential therapeutic activity already exists. In *S. aureus*, β-lactones have been found to inhibit Clp protease activity and decrease the virulence of the organism[36]. Additionally, the synergistic nature of ClpP1P2 protease depletion with aminoglycosides, a class of drugs already used to treat tuberculosis, points to a potential combination therapy against Mtb. As ClpP1P2 protease is most likely involved in preventing the accumulation of misfolded proteins and the degradation of critical endogenous regulatory proteins, small molecule modulators of ClpP1P2 activity would target a critical aspect of Mtb physiology, and might prove useful in the face of growing multi-drug resistance in one of the world's most successful pathogens.

MATERIALS AND METHODS

Ethics Statement

The animal experiments were preformed with protocols approved by the Harvard Medical School Animal Management Program, which is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, International (AAALAC) and meets National Institute of Health standards as set forth in the Guide for the Care and Use of Laboratory Animals (Revised, 2010). The institution also accepts as mandatory the PHS Policy on Humane Care and Use of Laboratory Animals by Awardee Institutions and NIH Principles for the Utilization and Care of Vertebrate Animals Testing, Research, and Training. An Animal Welfare Assurance of Compliance is on file with the Office of Laboratory Animal Welfare (OLAW) (#A3431-01).

Bacterial strains and plasmids

Msm mc^2155 (Msm) or Mtb H37Rv were grown at 37°C in Middlebrook 7H9 broth with 0.05% Tween 80 and ADC (0.5% BSA, 0.2% dextrose, 0.085% NaCl, 0.003 g catalase/1L media). Mtb

was additionally supplemented with oleic acid (0.006%). For growth curves, overnight cultures were diluted into the appropriate media and growth was either measured by OD_{600} or colony forming units per mL.

For complementation studies, wildtype Mtb ClpP1 and ClpP2 were amplified from H37Rv genomic DNA by PCR, using primers RMR01-RMR04, and ligated into the constitutively expressing plasmid pMV762zeo. C terminal 6X his or c-myc tags were added by PCR primers RMR05-RMR08 on Mtb ClpP1 and ClpP2 and recombined into the ATc inducible vector pTet using gateway recombination (Clontech). Site directed mutagenesis of ClpP1 and ClpP2 was carried out as described previously to generate various catalytic mutants used in the study. Catalytically inactive mutants were inserted into the ATc inducible vector pTet using gateway recombination. Processed Clp mutants were clonted into pTet or pMV762 vectors using primers listed above. The fusion GFP-SsrA was amplified from GFPmut3 wildtype DNA and cloned into pMV762zeo using primers RMR09-RMR12. Details of all plasmids and primers used in this study can be found below **Table 2.1** and **Table 2.2**, respectively.

Mycobacterial recombineering was employed, as described previously[21], to create strains ptet_clpP1P2 and clpP2_ID. For strain ptet_clpP1P2, the tetracycline promoter, tetracycline repressor, and a hygromycin resistance marker were inserted into p96863 (Genscript). Both upstream and downstream of the insertion site, p96863 contained 200 bp fragments flanking either side of the native clpP promoter. A linear PCR product containing the regions of homology, the hygromycin resistance marker, and tetracycline repressor and promoter was generated using primers RMR13 and RMR14. Allelic exchange of the native promoter was carried out by transformation of this linear substrate into a Msm strain expressing mycobacteriophage recombinases gp60 and gp61 on a nitrile inducible, counter-selectable episomal plasmid. Counter selection on 10% sucrose led to loss of the recombineering plasmid. Successful integration of the desired sequence was confirmed by PCR, using primers RMR13 and RMR16. As RMR16 lies

outside of the homology region used for recombineering, specific integration into the endogenous chromosome could be verified **(Figure 2.6)**.

To create strain clpP2_ID, a linear DNA substrate was created in a similar fashion. The inducible degradation tag was inserted into p54689 (Genscript). Both upstream and downstream of the insertion site, p54689 contained 200 bp of homology to the C terminal end of clpP2 and the 3'-UTR of clpP2, respectively. A linear PCR product containing this homology and the inducible degradation tag was generated using primers RMR15 and RMR16. This PCR product was transformed into Msm as described above. Successful integration of the desired sequence was confirmed by PCR, using primers RMR13 and RMR16. As RMR13 lies outside of the homology region used for recombineering, specific integration into the endogenous chromosome could be verified **(Figure 2.6)**.

To make the clpP2_ID strain, the tetracycline promoter and ClpP1 were inserted into the suicide plasmid, pSES. Constructs were electroporated into a Msm strain containing an integrated pMC1s vector constitutively expressing the tetR repressor[37]. Integrants were screened by PCR using primers RMR17 and RMR18.

Table 2.1 – Plasmids utilized for study

Plasmid	Properties/Uses
pTetOR::clpP1-myc	Inducible expression of c-myc-tagged Mtb ClpP1 (to assess in vivo interaction with ClpP2)
pTetOR::clpP2-his	Inducible expression of 6xHis-tagged Mtb ClpP2 (to assess in vivo interaction with ClpP1)
pTetOR::clpP1wt	Inducible expression of Mtb ClpP1 (for in vitro degradation assay)
pTetOR::clpP1S	Inducible expression of Mtb ClpP1-Ser^{98A} (for in vitro degradation assay, and overexpression in Mtb)
pTetOR::clpP2S	Inducible expression of Mtb ClpP2-Ser^{110A} (for in vitro degradation assay, and overexpression in Mtb)
pTetOR::clpP1H	Inducible expression of Mtb ClpP1-His^{123A} (for overexpression in Mtb)
p96863	Non-expressing, synthesized plasmid (Genscript) containing regions of homology to ClpP1 5'UTR and ORF. Used to generate linear PCR product for recombineering to create Msm ptet_clpP1P2
pKM339	Plasmid used to obtain tetracycline promoter, repressor, and hygromycin resistance marker, which were cut and inserted into p96863 for recombineering to create Msm ptet_clpP1P2
p54689	Non-expressing, synthesized plasmid (Genscript) containing regions of homology to ClpP2 ORF and 3'UTR. Used to generate linear PCR product for recombineering to create Msm clpP2_ID
puc57::inhA-ID	Plasmid used to obtain inducible degradation tag, which was cut and inserted into p54869 for recombineering to create Msm clpP2_ID
pSES::ptet_clpP1	Plasmid used for homologous recombination to create Msm ptet_clpP2
pMV762zeo::clpP1	Constitutively expressing plasmid expressing Mtb ClpP1 for complementation studies
pMV762zeo::clpP2	Constitutively expressing plasmid expressing Mtb ClpP2 for complementation studies
pMV762zeo::clpP1P2	Constitutively expressing plasmid expressing the entire clpP1clpP2 operon for complementation studies
pMV762zeo::GFP-SsrA	Constitutively expressing plasmid expressing the fusion construct GFP-SsrA to assess role of Clp protease in degradation of SsrA-tagged substrates
pSES::ClpP1	Plasmid used for homologous recombination event to create strain, pTet_clpP2
pNit::Che9c	Plasmid expressing mycobacteriophage recombinases, used in mycobacterial recombineering

Table 2.2 – Primers utilized for this study

Primer Name	Primer Sequence (5' to 3')	Use
RMR01	GCACTGTTAATTAAGAAGGAGATATACCT ATGCGTTCGAACTCGCAG	Cloning of processed Mtb ClpP1 into pTetOR, pmv (forward)
RMR02	AGTATACAGCTGTCACTGTGCTTCTCCAT TGACCTG	Cloning of processed Mtb ClpP1 into pTetOR (reverse)
RMR03	GCACTGTTAATTAAGAAGGAGATATACCT ATG CGCTACATCCTGCCGTC	Cloning of processed Mtb ClpP2 into pTetOR, pmv (forward)
RMR04	AGTATACAGCTGTCAGGCGGTTTGCGCG GA	Cloning of processed Mtb ClpP2 into pTetOR, pmv (reverse)
RMR05	AGTATACAGCTGTCACAGGTCCTCCTCC GAGATCAGCTTCTGCTCCTGTGCTTCTCC ATTGACCTG	Cloning of processed Mtb ClpP1-myc into pTetOR (reverse)
RMR06	AGTATACAGCTGGTGGTGGTGGTGGTGG TGCTGTGCTTCTCCATTGACCTG	Cloning of processed Mtb ClpP1-his into pTetOR (reverse)
RMR07	AGTATACAGCTGTCACAGGTCCTCCTCC GAGATCAGCTTCTGCTCGGCGGTTTGCG CGGA	Cloning of processed Mtb ClpP2-myc into pTetOR, pmv (reverse)
RMR08	AGTATACAGCTGGTGGTGGTGGTGGTGG TGGGCGGTTTGCGCGGA	Cloning of processed Mtb ClpP2-his into pTetOR, pmv (reverse)
RMR09	GATCCGCATGCTTAATTAAGAAGGAG	Cloning of GFP-ssrA into pmv (forward)
RMR10	GTGGTGGTGATGGATGGTGTTTGTATAG TTCATCCATGCCATG	Cloning of GFP-ssrA into pmv (reverse, first round)
RMR11	CTGATGTGAATCGGCGTGGTGGTGATGA TGGTGTTTCTATAG	Cloning of GFP-ssrA into pmv (reverse, second round to add first portion SsrA-tag)
RMR12	CGGAATATCGATCTAGGCAGCGAGAGCG TAGTCGCGCTGATGTGAATCGGC	Cloning of GFP-ssrA into pmv (reverse, third round to add remaining portion SsrA-tag)
RMR13	CCGCCGTGGCCTGACCATC	Generation of linear PCR product from p96863 to recombineer strain ptet_clpP1P2 (forward)
RMR14	TCTTCCGCCGACAGCAACAGG	Generation of linear PCR product from p96863 to recombineer strain ptet_clpP1P2 (reverse)
RMR15	CATCCAGGGCCAGTTCTC	Generation of linear PCR product from p54689 to recombineer strain clpP2_ID (forward)
RMR16	CGTGGTGTTTGCCGTTCT	Generation of linear PCR product from p54689 to recombineer strain clpP2_ID (reverse)
RMR17	GCACGGCATACATCATTTCGACGCCG	Used in screening for pTet_clpP2 strain (forward). Binds to the tetracycline promoter.
RMR18	GGCGGTTTGCGCGGAGAGC	Used in screening for pTet_clpP2 strain (reverse). Binds to the 3'-end of clpP2.

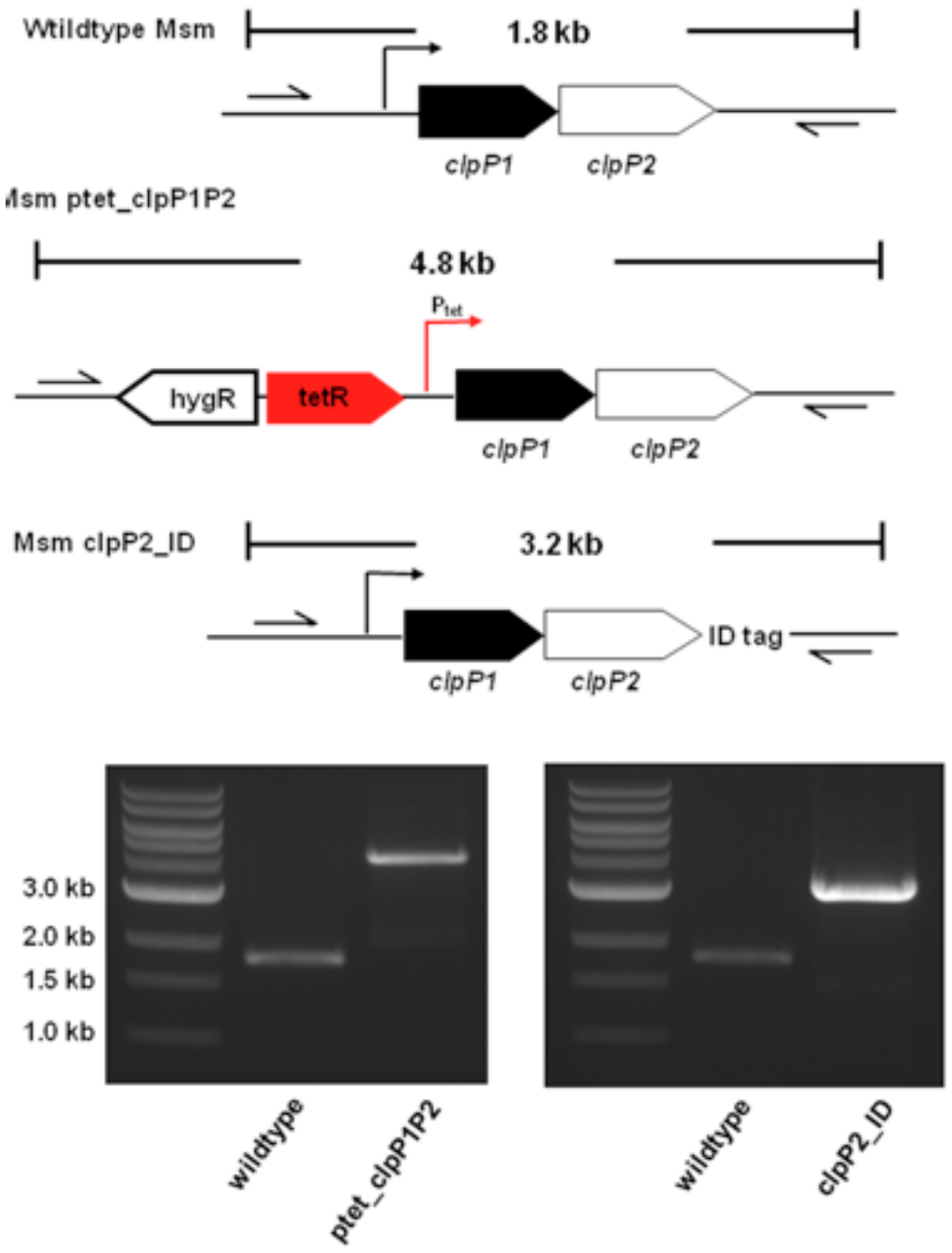


Figure 2.6 PCR confirmation of mycobacterial recombineering
Primers specific to the 5'-UTR and 3'-UTR (RMR13 and RMR16, arrows) were used to distinguish wildtype Msm (expected fragment: 1.8 kb), Msm ptet_ClpP1P2 (expected size: 4.8 kb), and Msm clpP2_ID (expected size 3.2 kb). For each construct, at least one primer was outside homology region used for recombineering in order to ensure specific insertion into the endogenous chromosome.

Inducible degradation of ClpP2

Mycobacterial recombineering was employed to insert the inducible degradation tag (ID-tag) directly downstream of the clpP2 open reading frame. Inducible degradation was performed as described previously. Briefly, strain clpP2_ID was transformed with an anhydrotetracycline inducible integrated plasmid carrying the HIV-2 protease. Stationary phase cultures were diluted 1000-fold, and induced with ATc (50 ng/mL). Cleavage by HIV-2 protease and subsequent protein degradation was monitored by epitope tags that flanked the ID-tag. To assess the role of the mycobacterial SsrA-tag, a constitutively expressing plasmid bearing GFP-SsrA was electroporated into clpP2_ID. Cultures were grown and induced as above, and increase in GFP upon HIV-2 protease induction was monitored by fluorescence (emission/excitation: 485/520) and by immunoblotting using anti-GFP.

Immunoblotting

Total protein lysates were prepared from equivalent cell numbers using bead beating and immunoblot analysis was performed. For figure 2.3 and 2.4, bacterial cultures were grown to from starting OD 0.05 for 3-18 hours in fresh medium with or without ATc. Primary antibodies anti-myc (Sigma Aldrich) and anti-FLAG (Sigma Aldrich) were used as specified by manufacturer. Detection was performed using secondary HRP conjugated antibodies (Pierce), and SuperSignal West Femto Chemiluminescent Substrate (ThermoScientific) according the manufacturer's protocol. In all cases, blots were stripped and reprobed with anti-RpoB (MyBiosource) to ensure equivalent loading of samples

Protein purification and in vitro peptidase assay

The C-terminally 6x His-tagged wild type *clpP1, wild type clpP2, clpP1Ser*^{98Ala}*,* and *clpP2Ser*^{110Ala} subunits were overexpressed in Msm using an anhydrotetracycline (ATc) inducible expression system. After overnight induction with ATc (100 ng/mL), cells were lysed by French press, and lysates were centrifuged for 1h at 100,000 g. The subunits were purified from the supernatant by

Ni-NTA affinity chromatography (Qiagen). Eluted fractions containing ClpP proteins were pooled and further purified by size exclusion chromatography on Sephacryl S-300 column. Equal amounts of ClpP1 and ClpP2 (1μg each) were mixed in the reaction buffer (50mM K-phosphate buffer pH 7,5, 100 mM KCl, 5% glycerol, 2 mM BME, 5 mM Z-Leu-Leu) and peptidase activity was measured by a rise in fluorescence at 460nm (Ex at 340 nm) with 0.1 mM Z-Gly-Gly-Leu-AMC as a substrate. To measure dominant negative effect of active site mutants, same reaction was carried out in the presence of 5 μg of the mutant proteins. Mass spectrometry was carried out by the Taplin Mass Spectrometry Facility.

Animal infections:

Six to eight week old C57BL/6 mice (Jackson Laboratory) were used for animal infections. Mice were infected via aerosolization with 5 x 10^6 CFU each of a 3:1 mixture of Mtb pTet::ClpP1 S98A and Mtb pTet::GFP (wild type Mtb transformed with a control pTet plasmid containing GFP). Mice were fed with chow with or without inducer doxycycline. At 30 days after infection, mice were sacrificed, lungs were homogenized and appropriate dilutions were plated on 7H10 plates containing hygromycin or kanamycin to select for the Clp mutant or the control respectively.

WORKS CITED

1. Ingmer, H. & Brøndsted, L. Proteases in bacterial pathogenesis. *Res Microbiol* **160**, 704–710 (2009).

2. Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899 (2003).

3. Glickman, M. H. & Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* **82**, 373–428 (2002).

4. Frees, D., Savijoki, K., Varmanen, P. & Ingmer, H. Clp ATPases and ClpP proteolytic complexes regulate vital biological processes in low GC, Gram-positive bacteria. *Molecular Microbiology* **63**, 1285–1295 (2007).

5. Burns, K. E., Pearce, M. J. & Darwin, K. H. Prokaryotic ubiquitin-like protein provides a two-part degron to Mycobacterium proteasome substrates. *J Bacteriol* **192**, 2933–2935 (2010).

6. Cerda-Maira, F. A. *et al.* Molecular analysis of the prokaryotic ubiquitin-like protein (Pup) conjugation pathway in Mycobacterium tuberculosis. *Molecular Microbiology* (2010).doi:10.1111/j.1365-2958.2010.07276.x

7. Darwin, K. H. The Proteasome of Mycobacterium tuberculosis Is Required for Resistance to Nitric Oxide. *Science* **302**, 1963–1966 (2003).

8. Katayama-Fujimura, Y., Gottesman, S. & Maurizi, M. R. A multiple-component, ATP-dependent protease from Escherichia coli. *J Biol Chem* **262**, 4477–4485 (1987).

9. Hwang, B. J., Park, W. J., Chung, C. H. & Goldberg, A. L. Escherichia coli contains a soluble ATP-dependent protease (Ti) distinct from protease La. *Proc Natl Acad Sci USA* **84**, 5550–5554 (1987).

10. Wang, J., Hartling, J. A. & Flanagan, J. M. The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* **91**, 447–456 (1997).

11. Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T. Linkage between ATP Consumption and Mechanical Unfolding during the Protein Processing Reactions of an AAA+ Degradation Machine. *Cell* **114**, 511–520 (2003).

12. Mogk, A. *et al.* Broad yet high substrate specificity: the challenge of AAA+ proteins. *Journal of Structural Biology* **146**, 90–98 (2004).

13. Farrell, C. M., Grossman, A. D. & Sauer, R. T. Cytoplasmic degradation of ssrA-tagged proteins. *Molecular Microbiology* **57**, 1750–1761 (2005).

14. Pruteanu, M. & Baker, T. A. Controlled degradation by ClpXP protease tunes the levels of the excision repair protein UvrA to the extent of DNA damage. *Molecular Microbiology* **71**, 912–924 (2009).

15. Gaillot, O., Bregenholt, S., Jaubert, F., Di Santo, J. P. & Berche, P. Stress-Induced ClpP Serine Protease of Listeria monocytogenes Is Essential for Induction of Listeriolysin O-Dependent Protective Immunity. *Infection and Immunity* **69**, 4938–4943 (2001).

16. Gaillot, O., Pellegrini, E., Bregenholt, S., Nair, S. & Berche, P. The ClpP serine protease is essential for the intracellular parasitism and virulence of Listeria monocytogenes. *Molecular Microbiology* **35**, 1286–1294 (2000).

17. Jenal, U. & Fuchs, T. An essential protease involved in bacterial cell-cycle control. *EMBO J* **17**, 5658–5669 (1998).

18. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology* **48**, 77–84 (2003).

19. Carroll, P., Faray-Kele, M.-C. & Parish, T. Identifying vulnerable pathways in Mycobacterium tuberculosis using a knock-down approach. *Applied and Environmental Microbiology* 1–19 (2011).doi:10.1128/AEM.02880-10

20. Akopian, T. *et al.* The active ClpP protease from M. tuberculosis is a complex composed of a heptameric ClpP1 and a ClpP2 ring. *EMBO J* 1–13 (2012).doi:10.1038/emboj.2012.5

21. van Kessel, J. C. & Hatfull, G. F. Mycobacterial recombineering. *Methods Mol Biol* **435**,

203–215 (2008).

22. Wei, J.-R. *et al.* Depletion of antibiotic targets has widely varying effects on growth. *Proc Natl Acad Sci USA* **108**, 4176–4181 (2011).

23. Frees, D. & Ingmer, H. ClpP participates in the degradation of misfolded protein in Lactococcus lactis. *Molecular Microbiology* **31**, 79–87 (1999).

24. Hahn, F. E., Wisseman, C. L. & Hopps, H. E. Mode of action of chloramphenicol. III. Action of chloramphenicol on bacterial energy metabolism. *J Bacteriol* **69**, 215–223 (1955).

25. Wyka, M. A. & St John, A. C. Effects of production of abnormal proteins on the rate of killing of Escherichia coli by streptomycin. *Antimicrob Agents Chemother* **34**, 534–538 (1990).

26. Smith, C. K., Baker, T. A. & Sauer, R. T. Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. *Proc Natl Acad Sci USA* **96**, 6678–6682 (1999).

27. Kiran, M. *et al.* Mycobacterium tuberculosis ftsH expression in response to stress and viability. *Tuberculosis (Edinb)* **89**, S70–S73 (2009).

28. Ingvarsson, H. *et al.* Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of Mycobacterium tuberculosis ClpP1. *Acta Crystallogr D Biol Crystallogr* **63**, 249–259 (2007).

29. Quon, K. C., Marczynski, G. T. & Shapiro, L. Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* **84**, 83–93 (1996).

30. Goldberg, A. L. Degradation of abnormal proteins in Escherichia coli (protein breakdown-protein structure-mistranslation-amino acid analogs-puromycin). *Proc Natl Acad Sci USA* **69**, 422–426 (1972).

31. Estorninho, M. *et al.* ClgR regulation of chaperone and protease systems is essential for Mycobacterium tuberculosis parasitism of the macrophage. *Microbiology* (2010).doi:10.1099/mic.0.042275-0

32. Sherrid, A. M., Rustad, T. R., Cangelosi, G. A. & Sherman, D. R. Characterization of a Clp Protease Gene Regulator and the Reaeration Response in Mycobacterium tuberculosis. *PLoS ONE* **5**, e11622 (2010).

33. Drag, M. & Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov* **9**, 690–701 (2010).

34. Kirstein, J. *et al.* The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. *EMBO Mol Med* **1**, 37–49 (2009).

35. Schmitt, E. K. *et al.* The Natural Product Cyclomarin Kills Mycobacterium Tuberculosis by Targeting the ClpC1 Subunit of the Caseinolytic Protease. *Angewandte Chemie (International ed in English)* (2011).doi:10.1002/anie.201101740

36. Böttcher, T. & Sieber, S. A. Beta-lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of Staphylococcus aureus. *J Am Chem Soc*

130, 14400–14401 (2008).

37. Ehrt, S. *et al.* Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res* **33**, e21 (2005).

Chapter 3:

Elucidating the mechanisms that underlie Clp protease essentiality in mycobacteria

Section 3.1 – Chapter Overview and Attributions

This chapter focuses on testing the hypothesis that the essentiality of Clp protease in mycobacteria is linked to the endogenous proteins that it degrades. In the most straightforward scenario, one could imagine that the importance of a degradative protease for growth could be explained by the enzyme's action on a toxic protein or repressor of normal growth. Coupling the generation of conditional clp mutants in Mtb and Msm with a robust isobaric quantitative proteomics methodology, we uncovered a set of proteins that were over-represented upon depletion of the protease. After systemic validation of the hits from proteomic screening, we identified an essential transcriptional repressor, WhiB1, as a likely Clp substrate. This finding was particularly interesting as blocking turnover of WhiB1 in a wildtype background inhibited growth and caused cell lysis, just as depletion of Clp protease. We were unable to confirm whether the essentiality of Clp was solely due to the turnover of WhiB1, but the potential protease substrates presented here (along with WhiB1) provide us with many avenues towards understanding the important biological processes that are regulated by Clp-dependent protein turnover.

Attributions. I composed the manuscript that forms this chapter, which we aim to submit shortly as a research article to Molecular Cell. Though I performed initial primer design, Jun-Rong Wei and Jessica T. Pinkham constructed the Mtb clpP1P2 depletion strain (P750-clpP1P2DAS). With this strain, I performed growth curve analysis and protein isolation for both immunoblotting to confirm Clp depletion and quantitative proteomics to identify potential substrates of Clp protease. I constructed the Msm clpP2 conditional mutant (clpP2_ID), and performed all experiments with the strain. I conducted isobaric multiplexed quantitative proteomics in the lab of Professor Steven Gygi under the mentorship of Mark Jedrychowski. While I prepped samples for proteomics and analyzed the resulting data with Mark's guidance, he graciously coordinated the running of samples on the instrumentation in the Gygi lab, and provided the detailed methods write-up for

MS analysis below. I performed all experiments related to validating proteins as likely Clp substrates, and probing the essentiality of WhiB1 turnover.

Section 3.2 – Proteomic profiling of proteins degraded by ClpP1P2 protease identifies WhiB1 as a substrate and provides insight into the mechanism of Clp essentiality in *Mycobacterium tuberculosis*

INTRODUCTION

Our understanding of how bacteria regulate cellular processes has long focused on the role of transcription factors in the activation or repression of responses, through modulation of mRNA production. In eukaryotes, however, elucidation of the ubiquitin-proteasome pathway has illustrated that targeted degradation of endogenous proteins is frequently employed as a regulatory mechanism[1,2]. Like eukaryotes, bacteria possess an array of compartmentalized proteolytic complexes, capable of degrading proteins into smaller polypeptides and amino acids. The main enzymes responsible for this in bacteria include Clp, FtsH, Lon, HslUV, the prokaryotic proteasome, and the HtrA family[3,4]. Initially, these complexes were thought to maintain protein quality control in their respective locales through the recognition of misfolded, mislocalized, or aberrant protein products. Several seminal studies expanded our understanding of these complexes by identifying an array of endogenous proteins that were targeted for degradation in bacteria[5,6]. While this suggested an important role of proteolysis in the regulation of bacterial physiology, it has been difficult to determine the functional significance of protein degradation by these proteolytic machines in bacteria, specifically distinguishing constitutive, passive recognition of a protein from coordinated, active proteolysis of a substrate as a regulatory mechanism.

Interestingly, *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis that kills nearly 1.3 million people annually[7], may provide a unique perspective into understanding the importance of targeted protein degradation in bacteria. In most model organisms, like *Escherichia*

coli and *Bacillus subtilis*, where the compartmentalized proteases have been extensively studied, they are mostly dispensable for normal growth[8,9]. However, a genome-wide screen for essential genes in Mtb suggested that numerous proteolytic complexes (namely Clp, FtsH, and HtrA) were all absolutely required for cell survival, providing evidence for their critical role in bacterial physiology[10]. It has been further shown, through the construction of conditional mutants, that Clp protease in mycobacteria is required for growth in vitro and survival in vivo in a mouse model of infection[11].

In mycobacteria, Clp is an ATP-dependent serine protease that is comprised of two stacked heptameric rings of ClpP1 and ClpP2 multimers, which are encoded in the same operon[12]. Each subunit possesses proteolytic activity, and the resulting heterotetradecameric structure is a compartmentalized protease with the fourteen active sites buried in an internal chamber. Based on homology to Clp in other organisms, it is hypothesized that proteins to be degraded enter through an axial pore that is regulated by the interaction of ClpP1P2 with various AAA+ ATPases (ClpC1 and ClpX in Mtb)[13,14]. These adapters have the dual function of recognizing substrates and providing the energy for protein unfolding and translocation into the proteolytic core. In Mtb, Clp has been implicated in the recycling of abnormal peptides stalled on the ribosome, through recognition of SsrA-tagged proteins, but elucidation of the endogenous proteins targeted by Clp has been lacking[11]. In general, the identification of substrates of essential proteases has not been systematically studied in bacteria. It is exactly in those cases, however, where a proteolytic complex is required for normal growth that understanding the proteins targeted for degradation will expand our understanding of how proteolysis is employed as a necessary regulatory mechanism.

In this study, we constructed a conditional Clp protease mutant in Mtb, and quantitatively compared the proteomes of Clp-deficient cells to cells containing wildtype levels of the enzyme. Because Clp is essential in mycobacteria, we hypothesized that degradation of a particular

protein (or set of proteins) was required for cell survival. We show here that Clp protease degrades WhiB1, an essential transcriptional repressor. Blocking Clp-dependent degradation of WhiB1 resulted in stabilization of WhiB1 in mycobacteria. Though still functional, this stabilized allele was toxic even at physiological levels, suggesting that proteolysis plays a critical role in regulating the amount of WhiB1 present. These data establish a potential mechanism for explaining the essentiality of Clp protease in mycobacteria, and provide some of the first evidence towards establishing the essential role that protein turnover plays in regulating mycobacterial physiology.

RESULTS

Clp protease is essential in *Mycobacterium tuberculosis*

The essentiality of Clp protease has been demonstrated in the non-pathogenic, fast growing model organism, *Mycobacterium smegmatis* (Msm), but not in *Mycobacterium tuberculosis* (Mtb). In order to do so, we constructed a Clp protease conditional mutant in Mtb that took advantage of complementary systems of promoter regulation and inducible protein degradation recently developed for use in mycobacteria **(FIGURE 3.1A)**. The strain was constructed by first introducing an additional copy of the *clpP1P2* operon into the Mtb chromosome at the attachment site of the phage L5, creating a *clpP1P2* merodiploid strain. This allele was modified to contain an anhydrotetracycline-responsive (ATc) promoter (P750) that would shut off transcription of the operon upon the addition of ATc (tet-off promoter)[15]. Additionally, the C-terminus of ClpP2 was modified to include a DAS tag. As described previously, the DAS tag facilitates targeted degradation of the tagged protein upon inducible expression of SspB[16]. Normally absent in mycobacteria, the expressed SspB recognizes the DAS tag and tethers the tagged protein to ClpXP for degradation. By regulating SspB expression with an ATc-inducible promoter (tet-on), the DAS inducible degradation system could be combined in the same strain with the tet-off promoter. Using mycobacterial recombineering[17], the endogenous *clpP1P2* operon was deleted,

creating a strain (P750-clpP1P2*DAS*) where *clpP1P2* expression could be repressed and circulating ClpP2-DAS could be degraded by the addition of ATc.

At low inoculums of $5X10^5$ CFU/mL addition of ATc (1.5 μg/mL) to P750-clpP1P2DAS had a bactericidal effect, demonstrating that ClpP1 and ClpP2 were also essential in Mtb **(FIGURE 3.1B)**. At higher inoculums, ie, $1X10^7$ CFU/mL, depletion also inhibited growth. These higher inoculums allowed us to harvest cellular material for protein and transcript expression analysis. Though the DAS inducible degradation system requires the ClpP1P2 proteolytic core for depletion of ClpP2, expression of SspB resulted in profound depletion of ClpP2-DAS within 48 hours, or two replicative cycles **(FIGURE 3.1C)**. Furthermore, qPCR analysis of *clpP1* and *clpP2* mRNA revealed that by 48 hours, transcription at the *clpP1P2* locus was significantly repressed in cultures exposed to ATc **(FIGURE 3.1D)**.

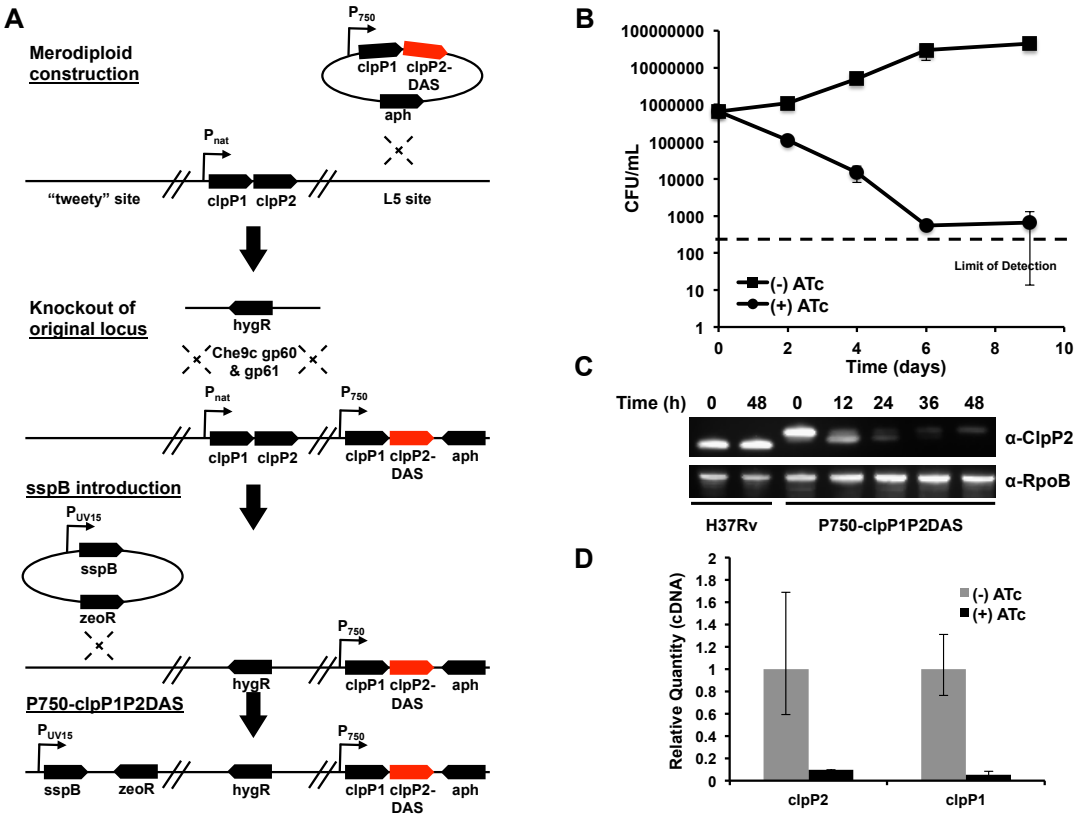


Figure 3.1 Clp protease is required for normal growth in *Mycobacterium tuberculosis*.

(A) Schematic depicting the construction of a conditional clpP1P2 mutant (Mtb P750-clpP1P2DAS). A stable, chromosomal merodiploid was first constructed by integrating an additional copy of the clpP1P2 operon at the L5 site. In this additional operon, the native promoter was replaced with a ATc responsive promoter, and the end of clpP2 was modified by appending a DAS inducible degradation motif. Next, the original clpP1P2 operon was deleted using mycobacterial recombineering. Lastly, a plasmid bearing ATc-inducible sspB was integrated into the tweety attB site on the chromosome, allowing for inducible degradation of ClpP2-DAS. (B) Growth curves of Mtb P750-clpP1P2DAS in the presence or absence of ATc (1.5 µg/mL). Data are represented as mean CFU/mL +/- standard deviation. (C) Depletion of ClpP2-DAS was tracked over time by Western blot in wt H37Rv and P750-clpP1P2DAS in the presence or absence of ATc (1.5 µg/mL). Blots were probed with α-ClpP2 and α-RpoB (loading control). (D) Quantitative PCR to determine clpP1 and clpP2 transcript levels using RNA generated from Mtb P750-clpP1P2DAS after growth for 48h in the presence or absence of ATc (1.5 µg/mL). Relative standard curves were generated for each probe set, and sigA transcript was used as an endogenous control. Data are represented as mean fold change, normalized to transcript in (-) ATc cultures +/- SEM of technical replicates.

Proteomic identification of Clp substrates in *Mycobacterium tuberculosis*

The P750-clpP1P2DAS strain afforded us a system to conditionally deplete the ClpP1P2 proteolytic core, and explore the mechanism of essentiality of Clp protease in Mtb. We hypothesized that growth inhibition observed in this strain resulted from an accumulation of Clp substrates that were either toxic to the cell or repressed normal growth processes. In order to determine this set of proteins, we turned to LC/MS3-based multiplexed isobaric quantitative proteomics. The use of isobaric tandem mass tags (TMT) has recently been used to simultaneously and robustly quantify protein levels in up to six conditions[18]. The six isobaric TMT tags (TMT126-TMT131) all bear the same weight, and so once the labeled peptides from the six experimental conditions are combined, the same peptides from each sample co-elute during fractionation, enter the mass spectrometer together and form a single peak in the initial mass spectrum. However, upon isolation and fragmentation of that single peak in a subsequent MS, the TMT reagent fragments. A differential distribution of isotopes on either side of the fragmentation site in the reagent results in a set of peaks between the mass range of 126-131 Da that reflects the relative ratios of that given peptide in a particular labeled, experimental condition.

In triplicate, we harvested P750-clpP1P2DAS grown up from an initial inoculum of $1X10^7$ CFU/mL for 48 hours either in the absence or presence of 1.5 μg/mL of ATc. Western blot analysis demonstrated a significant knockdown of ClpP2 in each of the cultures exposed to ATc **(FIGURE 3.2A)**, and so the six lysates were prepped for quantitative proteomics. Through TMT labeling followed by MS3-based quantitative proteomics, we were able to quantify 1564 Mtb proteins across the six conditions. Hierarchical clustering using Pearson correlational analysis[19] on the normalized, summed TMT intensities for each protein revealed a strong correlation between the three biological replicates of each condition **(FIGURE 3.2B)**. A total of 132 proteins were significantly over-represented in mut bacteria. We defined significant over-representation (or under-representation) as an average change of greater than or equal to (or less than or equal to) two-fold between mutant and wildtype conditions, and a p-value of less than or equal to 0.01

across the biological replicates **(FIGURE 3.2C, APPENDIX 2)**. Unfortunately, gene ontology (GO) analysis failed to reveal enrichment of any particular GO class among the over-represented proteins in Clp-depleted bacteria (data not shown). However, to support our theory tying Clp essentiality to the degradation of repressors of normal growth, there were numerous transcriptional modulators with increased abundance in mutant bacteria. Comparing protein intensity values between mutant and wildtype bacteria revealed a set of proteins that were highly over-represented (> 5 fold increase, 24 proteins), moderately over-represented (2-5 fold increase, 108 proteins), or under-represented (> 2 fold decrease, 23 proteins). Examples from each class are depicted in **FIGURE 3.2D**. Remarkably, ClpP1 and ClpP2 were the two most under-represented proteins in the screen, both depleted over 90% in mut cells compared to wt, thus confirming the efficiency of Clp depletion.

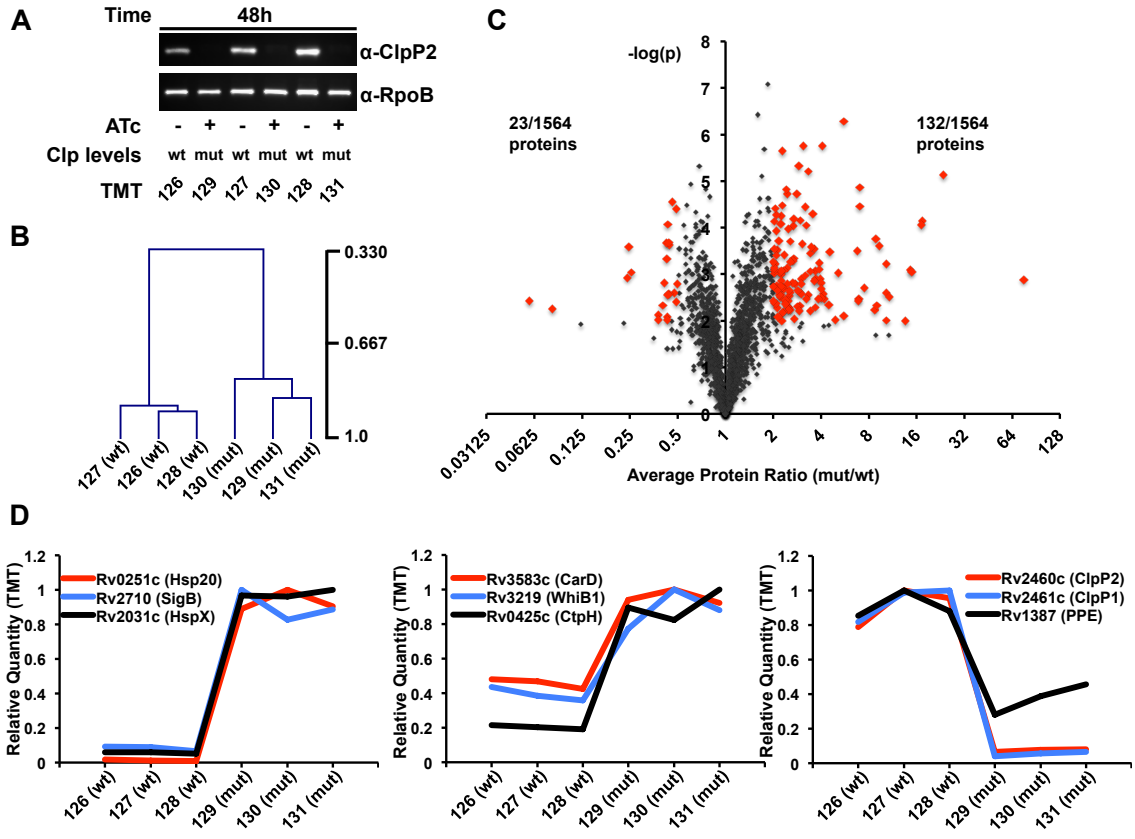


Figure 3.2 Proteomic profiling of P750-clpP1P2DAS in the presence and absence of ATc reveals a wide array of potential Clp protease substrates
(A) In triplicate, P750-clpP1P2DAS was grown for 48 hours in the absence, denoted "wt", or presence of ATc (1.5 µg/mL), denoted "mut", from a starting OD_{600} of 0.02. Western blotting of protein lysates with α-ClpP2 and α-RpoB (loading control) demonstrates degree of ClpP2-DAS depletion in mut cells. Samples were then used for TMT_6 MS3-based quantitative proteomics. The specific TMT reagent used for each condition is listed under the Western blot. (B) Normalized, summed intensities for all quantified proteins was used to perform Pearson correlational hierarchical clustering of biological replicates. (C) The Log2 ratio of average mutant protein intensity to average wildtype protein intensity plotted against the p-value determined by t-test, grouping the three biological replicates. The threshold for over-representation was set at an average ratio of greater than equal to 2, while the cut-off for under-representation was 0.5. In both instances, p-values below 0.01 were deemed significant. Proteins considered for further analysis are denoted in red. (D) The relative quantity of specific proteins plotted across the six TMT channels, for highly (left) and moderately (center) over-represented, and under-represented (right) proteins in the mutant versus wildtype conditions.

The transcriptional effectors WhiB1 and CarD are likely Clp protease substrates

There is only one reported Clp protease substrate in *Mycobacterium tuberculosis*, RseA, but unfortunately the anti-sigma factor was not quantified in our screen[20]. The lack of any other reported mycobacterial Clp substrates in the literature hindered our validation of the proteomic results. Furthermore, the growth inhibition that results from Clp protease depletion undoubtedly results in significant cellular stress and complicates the true cause of protein over-representation in the mutant cells. It is possible that an increase in abundance of any given protein in Clp-deficient bacteria could be due to a transcriptional stress response, and not due to stabilization from a lack of turnover. The presence of numerous heat shock proteins identified through our screening was suggestive of this. In order to determine the subset of over-represented proteins that were likely Clp substrates, we first turned to quantitative PCR analysis. Complementary DNA was generated from RNA isolated from P750-clpP1P2DAS grown for 48 hours in the presence or absence of ATc (1.5 μg/mL). Probe sets for candidates from the screen were used to compare levels of transcript between mutant and wildtype cells. This analysis revealed three groups, one where increases in protein abundance upon Clp depletion could be entirely explained by a transcriptional response clpS, asnC, dnaK, ndh, hsp20), a second where there was a clear discordance between protein abundance and transcript level changes (hspX, whib1, rpsL, merR), and a third where the difference was less clear (carD, papA3, hypothetical TF, luxR, sigB) **(FIGURE 3.4A)**. We posited that the latter two classes were more likely to contain Clp substrates, as the changes in abundance were more likely to be due to regulation at the protein level rather than transcriptional upregulation.

To further validate potential substrates of the Clp protease, we turned to a conditional clpP2 mutant (clpP2_ID Msm) we had previously developed in *Mycobacterium smegmatis* (Msm), a fast-growing, non-pathogenic model organism, closely related to Mtb. Briefly, this strain allowed for inducible degradation of ClpP2 upon the addition of ATc. Degradation was facilitated by an inducible degradation (ID) tag that was appended to the 3'-end of the chromosomal copy of clpP2.

The ID tag contained an SsrA tag that was masked by an HIV-2 protease recognition motif. Upon ATc-inducible expression of the HIV-2 protease, the motif was cleaved exposing the SsrA tag, and targeting ClpP2-ID for degradation. ClpP2 degradation could be tracked by epitope tags placed both upstream (myc) and downstream (FLAG) of the HIV-2 protease recognition site. We had previously reported that this mutant allowed for rapid degradation of ClpP2 (within 5 hours), and ClpP2 depletion resulted in the accumulation of a reporter substrate (GFP-SsrA), due to decreased turnover by Clp protease[11]. In earlier, unpublished work, we also used this strain to run a pilot TMT, MS2-based proteomic screen for potential Clp substrates. Using TMT six-plexing, we probed the proteomes of wildtype and clpP2_ID Msm inducibly expressing the HIV-2 protease at 0h, 5h, and 11h post addition of ATc. Western blot analysis revealed significant depletion of ClpP2-ID at both 5h and 11h **(FIGURE 3.3A)**.

In our proteomic screen, we quantified a total of 1776 Msm proteins across the six conditions, and hierarchical clustering using Pearson correlational analysis on the normalized, median TMT intensities for each protein appropriately revealed a strong correlation between the uninduced wildtype and clpP2_ID samples, the 5h and 11h induced wildtype samples, and two induced clpP2_ID samples **(FIGURE 3.3B)**. To minimize the probability of non-specific or stress-dependent protein over-representation, we used the 5h time point of both cultures for further analysis. Quantitative TMT analysis revealed that depletion of ClpP2 was nearly 2-fold at this time. A total of 107 proteins were over-represented in ClpP2-depleted cells at 5h, with over-representation defined as greater than or equal to two-fold increase in median protein intensity between clpP2_ID and wildtype samples. **(FIGURE 3.3C, APPENDIX 3)**.

Similar to our approach of validating potential substrates in Mtb, we conducted quantitative PCR analysis on RNA isolated from wildtype and clpP2_ID Msm cultures 5h post induction of HIV-2 protease. As in Mtb, this analysis revealed a class of targets that showed discordance between protein over-representation and changes in transcript levels between clpP2_ID and wildtype Msm

(FIGURE 3.3D). Interestingly, we identified two essential transcriptional effectors, CarD and WhiB1, with increased protein abundance in both Mtb and Msm proteomic screens. While whiB1 showed no significant fluctuations in transcript amount between Clp-deficient and Clp-replete cells in both Mtb and Msm, carD was discordant between Mtb and Msm qPCR experiments, with a slight increase in Mtb, but no significant change in Msm.

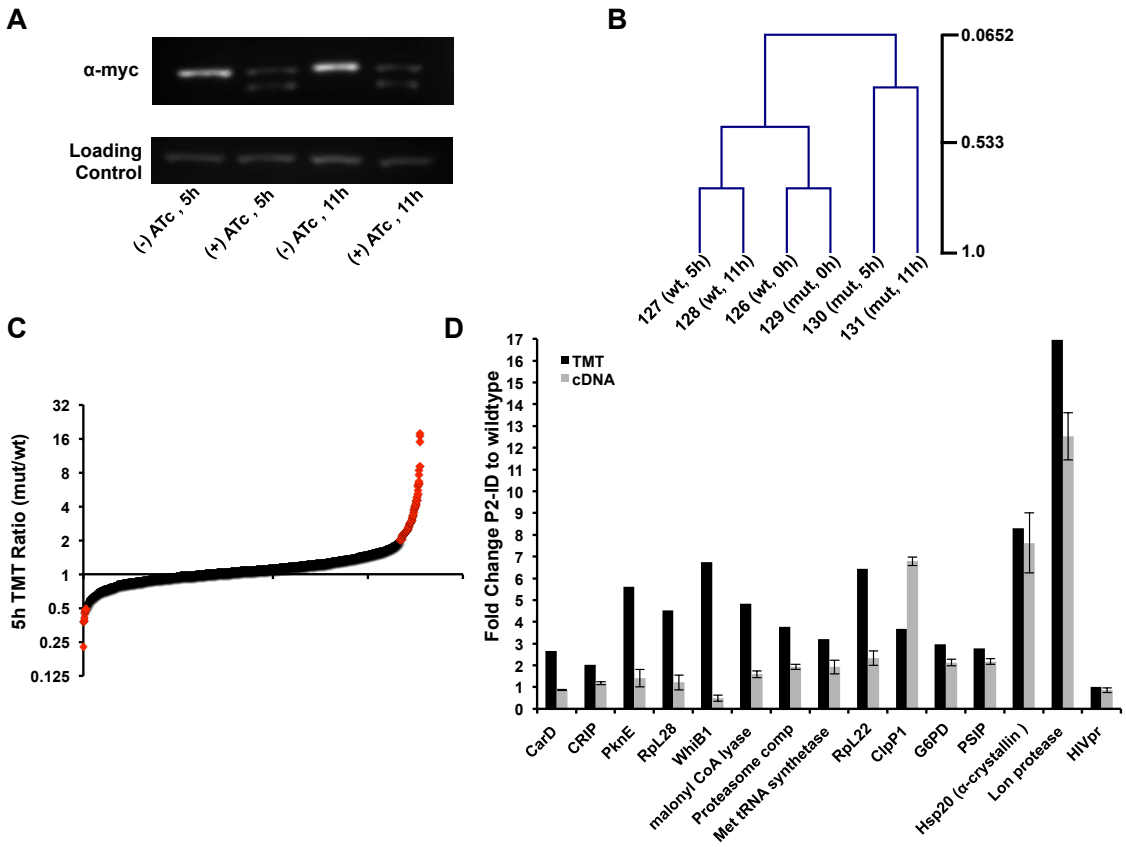


Figure 3.3 Proteomic profiling of clpP2-ID Msm in the presence and absence of ATc reveals a set of potential Clp protease substrates
(A) clpP2-ID Msm was grown in the presence or absence of ATc (100 ng/mL) from a starting OD_{600} of 0.04 for 5 or 11 hours. Depletion of ClpP2-ID was tracked by Western blot of protein lysates, probing for α-myc and α-RpoB (loading control). Samples were then used for TMT_6 MS2-based quantitative proteomics. (B) Normalized, median intensities for all quantified proteins was used to perform Perason correlational hierarchical clustering of the different conditions. The specific TMT reagent used for each condition is listed under the Western blot. (C) The Log2 ratios of median protein intensity at 5h for ClpP2 depleted cells (mut) to ClpP2 containing cells (wt). The threshold for over-representation was set at an average ratio of greater than or equal to 2, while the cut-off for under-representation was less than or equal to 0.5. Hits are denoted in red. (D) For a given set of proteins, the ratio of mutant to wildtype protein at 5 hours was compared to the ratio of transcript levels. Quantitative PCR was employed to determine transcript levels using RNA generated from clpP2-ID Msm after growth for 5h in the presence or absence of ATc (100 ng/mL). Relative standard curves were generated for each probe set, and sigA transcript was used as an endogenous control. For each target, data are represented as mean fold change, of mutant cells normalized to wildtype transcript amount +/- SEM of technical replicates.

As the clpP2_ID Msm strain appeared suitable for probing specific Clp recognition of endogenous proteins with measurable ClpP2 depletion within 5 hours, and CarD and WhiB1 are well conserved between Mtb and Msm, we used this strain to further validate the two proteins as Clp substrates. To provide complimentary evidence of WhiB1 and CarD degradation by Clp protease, we constructed GFP-fusion proteins, where GFP was appended to either the N- or C-terminus of each protein, and expressed these fusions in an episomal, ATc-inducible plasmid. By fusing GFP to either end of potential substrates, we hoped to track protein turnover via fluorescence while also determining whether the exposure of a particular terminus was important for protease recognition. Additionally, by tightly regulating the transcription of these constructs on an inducible plasmid, we hoped to prevent transcriptional modulation from confounding our findings. Expression in wildtype Msm revealed differential abundances, as measured by fluorescence, between N- and C-terminal fusions for each protein. For both WhiB1 and CarD, the N-terminal GFP fusions appeared to be less abundant than the C-terminal fusions **(FIGURE 3.4B, black bars)**. Despite the differential fluorescence between the two WhiB1 fusions, quantitative PCR analysis revealed that inducible expression of GFP-WhiB1 and WhiB1-GFP led to similar amounts of transcript in the cell, thus suggesting that any differential fluorescence observed was regulated at the protein level **(FIGURE 3.5)**.

To demonstrate that this discrepancy was specifically due to destabilization of one particular fusion by Clp protease, we introduced the fusions into clpP2_ID Msm, where addition of ATc simultaneously induced expression of each fusion construct and HIV-2 protease, and assessed protein abundances in the face of ClpP2 depletion. For both WhiB1 and CarD, depletion of ClpP2 resulted in an increase in the abundance of the N-terminal GFP fusion, relative to wildtype Msm **(FIGURE 3.4B)**, presumably reflecting stabilization due to reduced turnover. The differential abundance of the two fusions in wildtype Msm, and the selective stabilization of the N-terminal GFP fusion upon ClpP2 depletion, demonstrated that CarD and WhiB1 were likely substrates of Clp protease, that the Clp recognition motif for both was at the extreme C-terminus of the protein,

and that this motif was blocked upon fusion with GFP. Worry of transcriptional upregulation explaining the relative stabilization of the N-terminal GFP fusions upon ClpP2 depletion was mitigated as wildtype GFP revealed that transcription at the ATc responsive promoter was actually slightly repressed upon inducible degradation of the ClpP2 subunit (data not shown).

To ensure that destabilization was not a non-specific effect due to any non-specific N-terminal GFP fusions, we tested several other candidates using the same methodology. For both RpL28 and DnaA, C-terminal GFP fusions were actually less stable than their respective N-terminal constructs. In the case of RpL28, depletion of ClpP2 stabilized both fusions, suggesting that the motif for Clp recognition was internal and not dependent on an exposed terminus. DnaA fusions were not stabilized at all upon ClpP2 depletion suggesting that DnaA was either not a substrate or that both free ends were required for proteolysis **(FIGURE 3.6A)**.

Observing that turnover of WhiB1 required a free C-terminus, we hypothesized that WhiB1's degradation motif was at the extreme C-terminus. To show that the C-terminus was sufficient to confer destabilization and recognition by Clp protease, we constructed a variety of fusions where a variable number of C-terminal WhiB1 residues were appended to the end of GFP. Expressing these constructs on a constitutive episomal plasmid in clpP2_ID Msm revealed that the addition of the last fifteen, nine, and five amino acids of WhiB1 to GFP destabilized the protein with respect to wildtype GFP. Furthermore, wildtype levels of GFP protein expression were restored in these constructs upon ClpP2 depletion **(FIGURE 3.4C)**. Though not definitive determination of the specific motif, these experiments affirm that the C-terminus of WhiB1 enables Clp recognition, and that the terminal five amino acids are sufficient to destabilize GFP relative to the wildtype fluorescent protein. The WhiB1 degradation motif is most likely complex and spans several amino acids, as the different fusions appeared to progressively destabilize GFP as more C-terminal WhiB1 residues were appended. Additionally, we also observed that the C-terminal fifteen

residues of CarD destabilized GFP in wildtype Msm, relative to the wildtype fluorescent protein or the addition of the first fifteen amino acids of CarD to the N-terminus of GFP **(FIGURE 3.6B)**

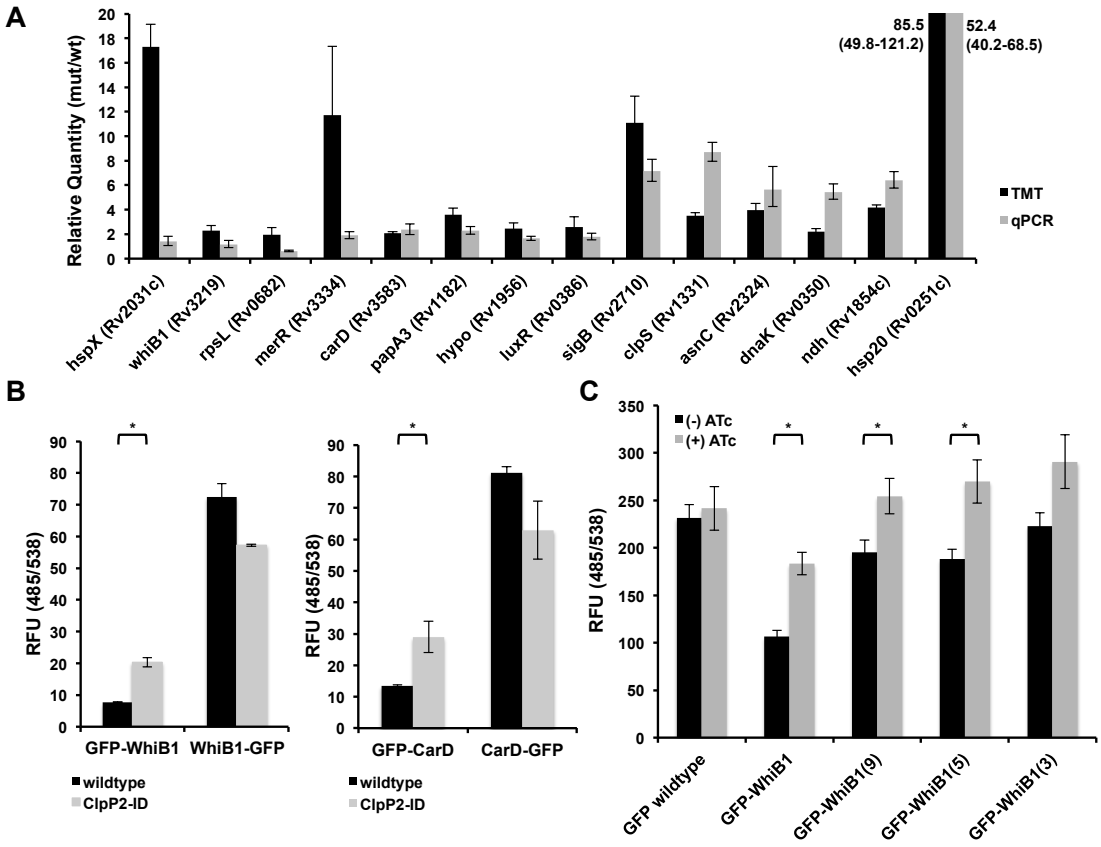


Figure 3.4 Validation of proteomic hits reveals that WhiB1 and CarD are likely Clp protease substrates

(A) Quantitative PCR (grey bars) to determine transcript levels of over-represented proteins upon depletion of Clp protease using RNA generated from Mtb P750-clpP1P2DAS after growth for 48h in the presence or absence of ATc (1.5 µg/mL). Relative standard curves were generated for each probe set, and sigA transcript was used as an endogenous control. Data are represented as mean fold change, normalized to transcript in (-) ATc cultures +/- SEM of technical replicates. Protein ratios (black bars) are derived from TMT experiments described in Figure 2, and represented as average ratios +/- standard deviation of biological replicates (B) Fluorescence (485/538) was measured for N- and C-terminal GFP fusions constructed for WhiB1 (left) and CarD (right), and induced for 8 hours in wildtype or clpP2-ID Msm with ATc (100 ng/mL). In clpP2-ID, ATc simultaneously induced fusion protein expression and degradation of ClpP2. (C) Fluorescence (485/538) was measured for GFP fusions bearing a variable number of C-terminal residues from WhiB1 in clpP2-ID Msm, grown in the absence or presence of ATc (100 ng/mL) for 8 hours. In (B) and (C), data are represented as mean RFU +/- standard deviation of biological replicates. Asterisks denote a p-value < 0.05 determined by t-test .

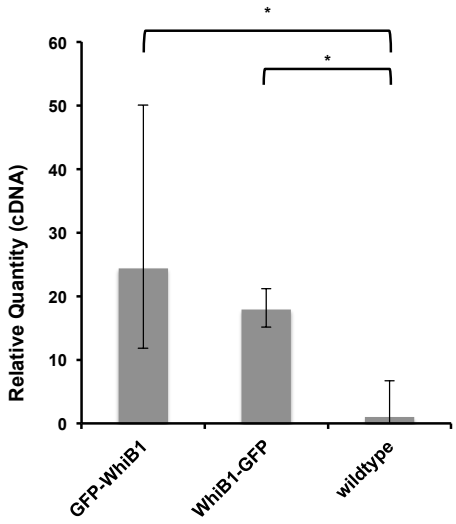


Figure 3.5 Overexpression of WhiB1 fusion constructs (GFP-WhiB1 and WhiB1-GFP) confirmed by quantitative PCR.
Quantitative PCR using a probe set that hybridized to both chromosomal and episomal copy of whiB1 to determine transcript abundance of whiB1 in strains inducibly over-expressing GFP-WhiB1 and WhiB1-GFP, compared to wildtype Msm. RNA was isolated from cultures grown for 6 hours from a starting OD_{600} of 0.06 in the presence of the inducer ATc (100 ng/mL). Relative standard curves were generated for each probe set, and sigA transcript was used as an endogenous control. Data are represented as mean fold change, normalized to transcript in wildtype cultures +/- SEM of technical replicates.

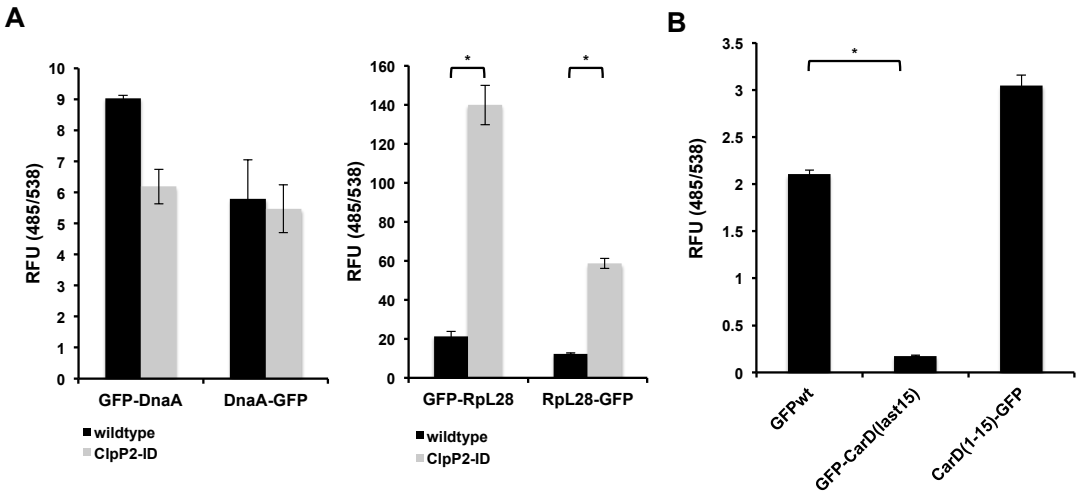


Figure 3.6 RpL28 is likely a Clp substrate, and the 15 terminal amino acids of CarD destabilize GFP relative to the wildtype fluorescent protein
(A) N- and C-terminal GFP-fusions for DnaA and RpL28, identified as potential Clp substrates from proteomic profiling of clpP2-ID Msm. Fluorescence (485/538) was measured for N- and C-terminal GFP fusions constructed for DnaA (left) and RpL28 (right), and induced for 8 hours in wildtype or clpP2-ID Msm with ATc (100 ng/mL). In clpP2-ID, ATc simultaneously induced fusion protein expression and degradation of ClpP2. (B) Fluorescence (485/538) was measured for wildtype GFP, and GFP fusions bearing either the N-terminal or C-terminal 15 amino acids from CarD expressed on a constitutive, episomal plasmid in wildtype Msm. In both (A) and (B), data are represented as mean RFU +/- standard deviation of biological replicates. Asterisks denote a p-value < 0.05 determined by t-test .

Stabilization of WhiB1, and blocking Clp-dependent degradation is toxic in mycobacteria

When performing experiments using the WhiB1 GFP fusion proteins, we made the interesting observation that prolonged over-expression of WhiB1-GFP inhibited the growth of wildtype Msm **(FIGURE 3.7A)**. This effect appeared to be specific to the C-terminal GFP fusion, as overexpression of the GFP-WhiB1 or even wildtype WhiB1 had no effect on growth. Through growth curve analysis, we were able to show that WhiB1-GFP was toxic and caused cell lysis **(FIGURE 3.7B)**.

Fusion proteins can often have non-specific toxic effects, and so we sought to determine if the WhiB1-GFP allele was still functional. As it has been previously shown that WhiB1 is capable of auto-repressing it's own transcription by binding to a site in the 5'utr of the gene[21], we determined transcript levels of the endogenous whiB1 locus upon overexpression of the two fusion proteins. Quantitative PCR revealed that transcription of the chromosomal whib1 5' utr was significantly repressed in the WhiB1-GFP compared to wildtype Msm, so the toxic WhiB1-GFP protein was still capable of binding DNA and modulating transcription **(FIGURE 3.7C)**. Unfortunately, qPCR analysis was unable to definitively determine whether the GFP-WhiB1 allele was functionally repressing whiB1 transcription. To rule out the possibility of differential toxicity being due to a non-functional N-terminal GFP fusion protein, we built a more sensitive reporter of promoter activity by fusing the 500 bp upstream of whiB1 to luciferase. By introducing this construct into strains inducibly expressing the GFP fusions, we could simultaneously monitor fluorescence for protein abundance and stability, and luminescence for whiB1 promoter activity. As expected, fluorescence measurements revealed that GFP-WhiB1 was more abundant than WhiB1-GFP. Both GFP-WhiB1 and WhiB1-GFP were able to repress transcription of whiB1 promoter compared to wildtype Msm, and the amount of repression appeared to correlate inversely with the amount of fusion protein present **(FIGURE 3.7D)**. The luciferase reporter demonstrates that both fusion proteins are comparably active and that the toxicity observed for WhiB1-GFP could be due

to increased WhiB1 abundance, a result of stabilization from lack of recognition of the blocked C-terminus by Clp protease.

In all of the above experiments, the fusion WhiB1 proteins were expressed on episomal, inducible plasmids. It is entirely feasible that the toxicity observed for WhiB1-GFP could be due to the dramatic overexpression resulting from the artificially constructed system. Furthermore, the increased abundance observed of WhiB1 in the proteomic screens was 2.2 fold in Mtb and 7 fold in Msm, much lower than the presumed over-expression from the inducible, episomal constructs. In order to test the functional effects of more physiological levels of the degradation-deficient WhiB1 protein, we constructed integrative plasmids in which the native whiB1 promoter regulated each fusion gene, and wildtype whiB1 as a control. Transformation of these plasmids into Msm resulted in significantly lower transformation efficiencies for the plasmid bearing WhiB1-GFP compared to those with GFP-WhiB1 and WhiB1wt **(FIGURE 3.7E)**. We believe the toxicity is even more severe than represented, as the WhiB1-GFP colonies that did grow were significantly smaller and took nearly twice as long to come up than the GFP-WhiB1 or control WhiB1wt transformations. The fitness cost associated with the introduction of a single copy of the fusion allele regulated by the native promoter demonstrates that slight disruption of WhiB1 turnover also effects growth. Furthermore, even though WhiB1 is able to auto-repress its own locus, the proteolysis of WhiB1 plays a dominant role in regulating the levels of WhiB1 and coordinating the functional effects of this presumed transcriptional repressor.

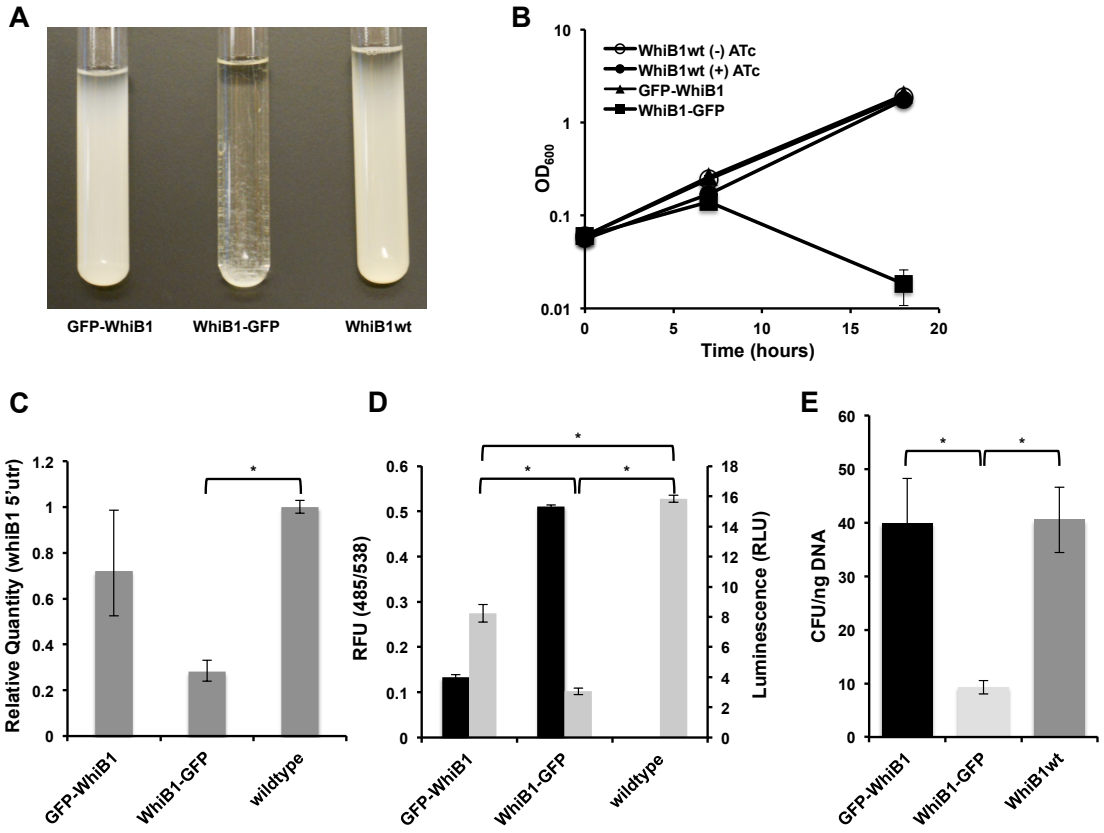


Figure 3.7 Blocking Clp-dependent degradation of WhiB1 is toxic
(A) Inducible expression of GFP-WhiB1, WhiB1-GFP, and WhiB1wt in Msm demonstrates growth inhibition of WhiB1-GFP expressing bacteria. Over-expression was induced with ATc (100 ng/mL). (B) Growth curves of strains expressing WhiB1 constructs in the presence of ATc (100 ng/mL). As a control, strains expressing WhiB1wt were grown in the absence of inducer. Data are represented as mean OD_{600} +/- standard deviation of biological replicates. (C) Quantitative PCR, using probe sets that hybridize to the whiB1 5'utr, to determine transcriptional repression at the endogenous whiB1 locus in wildtype Msm, or Msm inducibly expressing GFP-WhiB1 or WhiB1-GFP. RNA was isolated from cultures grown for 6 hours from a starting OD_{600} of 0.06 in the presence of the inducer ATc (100 ng/mL). Relative standard curves were generated for each probe set, and sigA transcript was used as an endogenous control. Data are represented as mean fold change, normalized to transcript in wildtype cultures +/- SEM of technical replicates. (D) Promoter reporters were constructed fusing the 500 bp upstream of whiB1 to luciferase. Luminescence (RLU, 100ms exposure, grey bars) was measured in wildtype Msm, or Msm inducibly expressing GFP-WhiB1 or WhiB1-GFP. Amounts of the fusion proteins were monitored by fluorescence (RFU, 485/538, black bars). Data are represented as RFU or RLU +/- standard deviation of biological replicates. (E) GFP whiB1 fusions and wildtype whiB1 were cloned into integrative plasmids, in which expression was driven by the native whiB1 promoter. Constructs were transformed into Msm, and transformation efficiency was determined by calculating colony forming units (CFU) obtained per ng DNA. In (C), (D), and (E), asterisks denote a p-value < 0.05, determined by t-test.

DISCUSSION

To date, only one endogenous substrate of Clp protease has been identified in mycobacteria. Global substrate identification of essential proteases has been a difficult enterprise. In model organisms where Clp is dispensable for normal growth, substrates were identified through pull-downs with a catalytically dead mutant that replaced the native protease[22]. In these cases, substrate identification was straightforward, but determining the functional importance of degradation of a particular protein was more challenging, due to the lack of clear impairment upon protease depletion. The essentiality of clpP1 and clpP2 thwarted multiple attempts to conduct allelic replacement pull-downs in mycobacteria, and so we employed a recently developed method of conditional gene expression for mycobacteria to construct an Mtb strain where the levels of ClpP1P2 protease could be tightly regulated. In addition to confirming the essentiality of Clp protease in Mtb, we coupled the generation of this strain to a new robust methodology of quantitative proteomics to identify a set of proteins in both Mtb and Msm that were over-represented in protease deficient cells. The TMT-based proteomics method we employed has been shown to reliably determine two-fold differences in the protein abundance between samples[23].

Because protein over-representation in mutant bacteria did not require a direct interaction with Clp, proteins identified from this screening method could have been over-represented for a multitude of reasons, including transcriptional upregulation in response to the stress of Clp depletion. Therefore, the candidates obtained from our proteomics experiments were only seen as potential substrates. Through a combination of mRNA transcript analysis, to show that protein over-representation in Clp-deficient cells was not due to increases in transcript abundance, and the creation of whole protein N-terminal and C-terminal GFP fusions, to monitor protein stability and turnover, we isolated two essential transcriptional modulators, WhiB1 and CarD, as likely Clp protease substrates. Though limited by the lack of an in vitro assay for testing the degradation of whole proteins by the ClpC1P1P2 or ClpXP1P2 complexes, we employed complimentary fusion

protein approaches and showed that the isolated C-termini of WhiB1 and CarD were sufficient to destabilize GFP in a Clp-dependent manner.

While there are likely other substrates identified in our screen, we chose to focus on WhiB1 in this study due to the interesting observation that expression of WhiB1-GFP was toxic to mycobacteria, while GFP-WhiB1 was not. Having generated evidence that C-terminal fusion of GFP blocked degradation by Clp, we believed that the toxicity of WhiB1-GFP may be linked to lack of proteolysis. Further evidence of this hypothesis was provided by the fact that the fusion allele was functional, and capable of binding DNA to repress transcription. While there is a chance that WhiB1-GFP toxicity is still non-specific, we believe that we have mitigated these concerns by showing the allele is still functional, and that it is associated with a fitness cost at physiological levels. Given that WhiB1 can auto-repress its own locus, and toxicity still results when a native promoter governs the fusion, we believe that tightly regulated turnover of WhiB1 is critical for normal growth.

Why could the turnover of WhiB1 be required for growth? In mycobacteria, there is substantial controversy regarding the role of the WhiB proteins, though most agree that they play an important role in redox biology. WhiB proteins are capable of binding redox-sensitive [4Fe-4S] clusters, which can serve as redox-active co-factors or a switches that reflect the reductive and oxidative potential of a cell[24]. Some studies have provided evidence that the WhiB proteins are disulfide reductases[25], but the prevailing thought is that this family is comprised of transcription factors capable of modulating cellular processes that are intimately tied to the redox state of the cell or perceived oxidative stress[26]. In Mtb, WhiB1 was found to be an essential DNA binding protein capable of auto-repressing it's own transcription[21]. Beyond this, the WhiB1 regulon or functional effects of WhiB1 DNA binding have yet to be definitively elucidated. As blocking Clp recognition of WhiB1 stabilizes the repressor, it would be a logical extension to assume that the elevated levels of the transcription factor contributes to the lethality of Clp deficient cells by

repressing essential genes. Interestingly, ChIP-Seq to determine the WhiB1 regulon in Mtb has been undertaken, and preliminary data suggests the presence of 71 binding sites for WhiB1, thirteen of which are essential[27]. For example, some genes in the putative WhiB1 regulon are involved in the synthesis of metabolites, such as heme and riboflavin. Insufficient amounts of these important co-factors resulting from lack of WhiB1 turnover may explain the lethality of degradation-deficient WhiB1, and in turn the essentiality of Clp protease. Alternatively, extremely elevated levels of WhiB1, due to absence of turnover, may result in non-canonical binding of the transcription factor to other sites not determined by ChIP-Seq methods. In this case, the lethality of WhiB1 may result from non-specific repression of other essential genes, and ChIP-Seq on a stabilized form of WhiB1 or RNA-Seq would need to be done in order to determine why lack of proteolysis is lethal.

While both stabilization of WhiB1-GFP and Clp depletion are lethal, we can only suggest that turnover of WhiB1 is the major critical role of Clp protease in mycobacteria. In reality, however, it is likely that the mechanism of Clp essentiality is multifactorial. It has recently been found that the tmRNA pathway in mycobacteria is critically required for normal growth (Zhang YJ, manuscript submitted). In this pathway, peptides stalled on the ribosome are modified through trans-translational addition of an SsrA tag, released from the ribosome, and subsequently degraded by Clp protease[28]. The tmRNA pathway is not essential in other organisms, and the reasons for the divergent essentiality of the tmRNA pathway in mycobacteria are unknown. It is likely that another critical role of Clp protease in cellular physiology is to serve as the effector of SsrA tagging. Additionally, genome wide screening suggests that both ATPases that interact with the ClpP1P2 proteolytic core, ClpC1 and ClpX, are essential in Mtb[10]. The fact that both these adapters are required supports the hypothesis that there are multiple substrates that need to be recognized, at least one per ATPase, by Clp protease to facilitate normal growth. The WhiB1 story highlights one critically important substrate of Clp protease in mycobacteria, but there are likely several others whose stabilization may have similar or more subtle effects than the toxic effect observed

upon blocking WhiB1 turnover. Though overexpression of degradation-deficient alleles of CarD did not explicitly inhibit growth, it is interesting to note that CarD has been implicated in the stringent response, and directly interacts with the beta-subunit of the RNA polymerase to down regulate transcription of the translational machinery and amino acid biosynthetic enzymes[29]. While stabilization of CarD alone is not sufficient to cause toxicity in mycobacteria, the transcriptional repression of enzymes important for vegetative growth facilitated by CarD may contribute to the growth inhibition observed upon depletion of Clp protease, and partially explain the essentiality of Clp protease in mycobacteria.

The demonstration that turnover of WhiB1 is specifically required for bacterial viability provides unique insight into the active role that targeted protein degradation plays in coordinating cellular physiology. WhiB1 is thought to be a transcription factor capable of auto-repression. However, even when the native promoter regulates a degradation deficient allele of WhiB1 at physiological levels, a fitness cost is associated with the presence of the proteolytically-dysregulated allele. This suggests that proteolysis plays a dominant role in regulating the functional impact of WhiB1 by coordinating the total amount of WhiB1 in the cell. This is especially intriguing considering that WhiB1 is essential in mycobacteria and deletion of the gene also prevents normal growth. Thus, the levels of WhiB1 in the cell must be tightly titrated, and despite WhiB1's ability to repress it's own locus, lack of WhiB1 proteolysis results in the presence of an aberrant amount of the transcription factor. By demonstrating that proteolysis of a particular protein can be co-dominant with transcriptional modulation, we provide evidence that the essential Clp protease in mycobacteria is employed to actively coordinate critical aspects of cellular physiology.

To date, there are only a few examples in bacteria where targeted proteolysis has been proven to be the dominant player in regulating a vital cellular process. In *Caulobacter crescentus*, one of the few bacteria where Clp protease is also essential, degradation of CtrA by Clp is absolutely necessary for the transition of a non-replicating, motile swarmer cell into a replicating, stalked

cell[30]. In most other cases, proteolysis of specific endogenous substrates has been confirmed, but proving that targeted turnover is the main mode of regulation of that protein or the pathway that the substrate is involved in has proven difficult. Here, we show that proteolysis of WhiB1 by Clp protease is required for growth, and absence of this turnover has a profound effect on the viability of mycobacteria.

In summary, we have highlighted a potential list of substrates of the essential Clp protease in mycobacteria, and provided a framework for confirming degradation and probing the functional significance of targeted turnover of endogenous proteins by essential proteases. While our work with WhiB1 highlights the importance of proteolysis in maintaining normal cellular physiology for one particular protein, the identification of other critical substrates will move us towards a more holistic understanding of how bacteria coordinate cellular activities through the integration of transcriptional and proteolytic regulation, both of which are likely equally important in maintaining bacterial viability.

MATERIALS AND METHODS

Bacterial strains and plasmids

Msm mc^2155 (Msm) or Mtb H37Rv were grown at 37°C in Middlebrook 7H9 broth with 0.05% Tween 80 and ADC (0.5% BSA, 0.2% dextrose, 0.085% NaCl, 0.003 g catalase/1L media). Mtb was additionally supplemented with oleic acid (0.006%). For growth curves, overnight cultures were diluted into the appropriate media and growth was either measured by OD_{600} or colony forming units per mL.

A complete list of plasmids and primers used for cloning this study can be found in **TABLE 3.1** and **TABLE 3.2**. The construction of strain clpP2_ID Msm has been described in depth in Chapter 2.

TABLE 3.1 – Plasmids utilized in this study

Plasmid	Properties/Uses
pNIT(zeo)::RecET-SacB	Nitrile inducible plasmid to make Mtb competent for recombineering
pEXPR(kan)::RT38-p750-clpP1P2DAS	L5 integrating plasmid used to construct clpP1P2 merodiploid Mtb
pGMCtKq22(zeo)::TSC10M1-pUV15-sspB	ATc-inducible expression of sspB for DAS inducible degradation of ClpP2-DAS
pTetOR(zeo)::gfp-attR	Gateway destination vector for N-terminal GFP tagging
pTetOR(zeo)::attR-gfp	Gateway destination vector for C-terminal GFP tagging
pTetOR(zeo)::gfp-carD	ATc-inducible expression of GFP-CarD
pTetOR(zeo)::gfp-whiB1	ATc-inducible expression of GFP-WhiB1
pTetOR(zeo)::gfp-rpL28	ATc-inducible expression of GFP-RpL28
pTetOR(zeo)::gfp-dnaA	ATc-inducible expression of GFP-DnaA
pTetOR(zeo)::carD-gfp	ATc-inducible expression of CarD-GFP
pTetOR(zeo)::whiB1-gfp	ATc-inducible expression of WhiB1-GFP
pTetOR(zeo)::rpL28-gfp	ATc-inducible expression of RpL28-GFP
pTetOR(zeo)::dnaA-gfp	ATc-inducible expression of DnaA-GFP
pMV762(zeo)::gfp-whiB1(last15)	Constitutive expression of GFP-WhiB1(last 15 amino acids)
pMV762(zeo)::gfp-whiB1(last9)	Constitutive expression of GFP-WhiB1(last 9 amino acids)
pMV762(zeo)::gfp-whiB1(last5)	Constitutive expression of GFP-WhiB1(last 5 amino acids)
pMV762(zeo)::gfp-whiB1(last3)	Constitutive expression of GFP-WhiB1(last 3 amino acids)
pMV762(zeo)::gfp-carD(last15)	Constitutive expression of GFP-CarD(last 15 amino acids)
pGH1000A(hyg)::PwhiB1(500)-Luciferase	Integrative luciferase reporter to monitor whiB1 promoter activity

TABLE 3.2 – Primers utilized in this study

Primer Name	Primer Sequence (5' to 3')	Use
attB4-RT38F	GGGGACAACTTTGTATAGAAAAGTTG CAGCTGGCTAGCGAGTCATG	Cloning of RT38 into multisite gateway destination vector (forward)
attB1-RT38R	GGGGACTGCTTTTTTGTACAAACTTG AATATTGGATCACGCCGCGA	Cloning of RT38 into multisite gateway destination vector (reverse)
attB1-p750F	GGGGACAAGTTTGTACAAAAAAGCAG GCTGCTACCAGGCCTAGATCTGG	Cloning of p750 into multiple gateway destination vector (forward)
attB2-p750R	GGGGACCACTTTGTACAAGAAAGCTG GGTGGTGGTGCATGCGGTTGTGA	Cloning of p750 into multiple gateway destination vector (reverse)
attB2-clpP1P2DASF	GGGGACAGCTTTCTTGTACAAAGTGG GAAGGAGATATACCTGTGAGCCAAGT GACTGAC	Cloning of clpP1P2DAS into multiple gateway destination vector (forward)
attB3-DASR	GGGGACAACTTGTATAATAAAGTTGT CACTAGCTGGCGTCCGCGTAGTTCTC GGAGTAGT	Cloning of clpP1P2DAS into multiple gateway destination vector (round 2, reverse)
clpP1P2DASR	GTAGTTCTCGGAGTAGTTCTCGTCGT TGGCGGCGGCGGTTTGCGCGGAGAG CTTC	Cloning of clpP1P2DAS into multiple gateway destination vector (round 1, reverse)
5flankF	GCTCGTCACGCGCGGCGTT	Cloning of upstream fragment for stitch PCR to generate recombineering product to delete clpP1P2 operon (forward)
5flankR	GGGGAGTATAACTTGGGGCACCTGCT TTCCTCGA	Cloning of upstream fragment for stitch PCR to generate recombineering product to delete clpP1P2 operon (reverse)
hygkoF	AAGCAGGTGCCCCAAGTTATACTCCC CGACGTGGCC	Cloning of hygR for stitch PCR to generate recombineering product to delete clpP1P2 operon (forward)
hygkoR	TCAGGCGGTTTGCGTCTAGACTCGAG GTACCGGCG	Cloning of hygR for stitch PCR to generate recombineering product to delete clpP1P2 operon (reverse)
3flankF	CTCGAGTCTAGACGCAAACCGCCTGA GCCATGG	Cloning of downstream fragment for stitch PCR to generate recombineering product to delete clpP1P2 operon (forward)
3flankR	ATGCTCAATCTGCAGCGGTCGC	Cloning of downstream fragment for stitch PCR to generate recombineering product to delete clpP1P2 operon (reverse)
clpPkoF	GGAAGCTCAGGTTACCGTCA	Primer to check deletion of clpP1P2 operon in Mtb (forward)
hygR	GCGTAGGAATCATCCGAATC	Primer to check deletion of clpP1P2 operon in Mtb (forward)
RMR257	AAACCCTTAATTAAGAAGGAGATATA CCTATGGCTA	Cloning of gfp for stitch PCR to generate N-terminal GFP Gateway tagging insert (forward)
RMR258	ACAAACTTGTTTTGTATAGTTCATCCA TGCCATG	Cloning of gfp for stitch PCR to generate N-terminal GFP Gateway tagging insert (reverse)

TABLE 3.2 (continued) – Primers utilized in this study

RMR259	AACTATACAAAACAAGTTTGTACAAAA AAGCT	Cloning of attB ccdB-CM^R cassette for stitch PCR to generate N-terminal GFP Gateway tagging insert (forward)
RMR260	TTTTTTGATATCACCACTTTGTACAAG AAAGCTGAAC	Cloning of attB ccdB-CM^R cassette for stitch PCR to generate N-terminal GFP Gateway tagging insert (reverse)
RMR261	AAACCCTTAATTAAACAAGTTTGTACA AAAAAGCT	Cloning of attB ccdB-CM^R cassette for stitch PCR to generate C-terminal GFP Gateway tagging insert (forward)
RMR262	CCTTTGCTAGCAACCACTTTGTACAA GAAAGCTGAAC	Cloning of attB ccdB-CM^R cassette for stitch PCR to generate C-terminal GFP Gateway tagging insert (reverse)
RMR263	CAAAGTGGTTGCTAGCAAAGGAGAAG AACTT	Cloning of gfp for stitch PCR to generate C-terminal GFP Gateway tagging insert (forward)
RMR264	TTTTTTGATATCTCATTTGTATAGTTCA TCCATGCCAT	Cloning of gfp for stitch PCR to generate C-terminal GFP Gateway tagging insert (reverse)
RMR271	GGGGACAAGTTTGTACAAAAAAGTTG CCCATATTTTCAAGGTCGGAGACACC GT	attB Gateway carD PCR product for N-terminal GFP tagging (forward)
RMR272	GGGGACCACTTTGTACAAGAAAGCTG GGTCTCAAGACGCGGCGGCTAAAAC	attB Gateway carD PCR product for N-terminal GFP tagging (reverse)
RMR273	GGGGACAAGTTTGTACAAAAAAGTTG CCCATGAAGGAGATATACCTATGATT TTCAAGGTCGGAGACAC	attB Gateway carD PCR product for C-terminal GFP tagging (forward)
RMR274	GGGGACCACTTTGTACAAGAAAGCTG GGTCAGACGCGGCGGCTAAAACCTC	attB Gateway carD PCR product for C-terminal GFP tagging (reverse)
RMR275	GGGGACAAGTTTGTACAAAAAAGTTG CCCATGATTGGCGCCACAAGGCGGT	attB Gateway whiB1 PCR product for N-terminal GFP tagging (forward)
RMR276	GGGGACCACTTTGTACAAGAAAGCTG GGTCTCAGACCCCGGTACGGGCTTTC	attB Gateway whiB1 PCR product for N-terminal GFP tagging (reverse)
RMR277	GGGGACAAGTTTGTACAAAAAAGTTG CCCATGAAGGAGATATACCTATGGAT TGGCGCCACAAGGC	attB Gateway whiB1 PCR product for C-terminal GFP tagging (forward)
RMR278	GGGGACCACTTTGTACAAGAAAGCTG GGTCGACCCCGGTACGGGCTTTCG	attB Gateway whiB1 PCR product for C-terminal GFP tagging (reverse)
RMR279	GGGGACAAGTTTGTACAAAAAAGTTG CCCATTCCGCCCACTGCCAAGTCAC	attB Gateway rpL28 PCR product for N-terminal GFP tagging (forward)
RMR280	GGGGACCACTTTGTACAAGAAAGCTG GGTCTCAGATCCGCTGCCCCTGG	attB Gateway rpL28 PCR product for N-terminal GFP tagging (reverse)
RMR281	GGGGACAAGTTTGTACAAAAAAGTTG CCCATGAAGGAGATATACCTTTGTCC GCCCACTGCCAAGT	attB Gateway rpL28 PCR product for C-terminal GFP tagging (forward)
RMR282	GGGGACCACTTTGTACAAGAAAGCTG GGTCGATCCGCTGCCCCTGGCGA	attB Gateway rpL28 PCR product for C-terminal GFP tagging (reverse)
RMR283	GGGGACAAGTTTGTACAAAAAAGTTG CCCATACCGATGACCCCGGTTCAGGC	attB Gateway dnaA PCR product for N-terminal GFP tagging (forward)

TABLE 3.2 (continued) – Primers utilized in this study

RMR284	GGGGACCACTTTGTACAAGAAAGCTG GGTCCTAGCGCTTGGAGCGCTGAC	attB Gateway dnaA PCR product for N-terminal GFP tagging (reverse)
RMR285	GGGGACAAGTTTGTACAAAAAAGTTG CCCATGAAGGAGATATACCTTTGACC GATGACCCCGGTTCA	attB Gateway dnaA PCR product for C-terminal GFP tagging (forward)
RMR286	GGGGACCACTTTGTACAAGAAAGCTG GGTCGCGCTTGGAGCGCTGACGG	attB Gateway dnaA PCR product for C-terminal GFP tagging (reverse)
RMR248	AAAAAAATTAATTAAGAAGGAGATATA CCTATGGCTA	Cloning of gfp truncations with variable number of 3' base pairs from whiB1 and carD (forward)
RMR249	AAAAAAAATCGATTCAGACCCCGGTA CGGGCTTTCGTGCGGGCGTTGCGAC GCTTCAGTTTGTATAGTTCATCCATGC CA	Cloning of gfp-whiB1[last 15 amino acids] (reverse)
RMR294	AAAAAAAATCGATTCAGACCCCGGTA CGGGCTTTCGTGCGGGCTTTGTATAG TTCATCCATGCCA	Cloning of gfp-whiB1[last 9 amino acids] (reverse)
RMR295	AAAAAAAATCGATTCAGACCCCGGTA CGGGCTTTGTATAGTTCATCCATGCC A	Cloning of gfp-whiB1[last 5 amino acids] (reverse)
RMR296	AAAAAAAATCGATTCAGACCCCGGTT TTGTATAGTTCATCCATGCCA	Cloning of gfp-whiB1[last 3 amino acids] (reverse)
RMR298	AAAAAAAATCGATTCAAGACGCGGCG GCTAAAACCTCGTCAAGGATGGTCTC GGCTTTGGCTTTGTATAGTTCATCCAT GCCA	Cloning of gfp-carD[last 15 amino acids] (reverse)
RMR313	AAAAAAATCTAGAGGGGTGTTTGCGA CGACCAG	Cloning of PwhiB1(500 bp) for stitch PCR to create luciferase promoter reporter (forward)
RMR310	GGCGTCTTCCATGTGATCTAACTCCT AATCGGGCGC	Cloning of PwhiB1(500 bp) for stitch PCR to create luciferase promoter reporter (reverse)
RMR311	AGGAGTTAGATCACATGGAAGACGCC AAAAACATAAAGAAAGG	Cloning of luciferase for stitch PCR to create luciferase promoter reporter (reverse)
RMR312	AAAAAAAAAGCTTCTATTTCACGGCG ATCTTTCCGC	Cloning of luciferase for stitch PCR to create luciferase promoter reporter (reverse)

The first step in the construction of P750-clpP1P2DAS was transformation of the attB L5 site integrating plasmid, pEXPR(kan)::RT38-p750-P1P2DAS into H37Rv/pNIT(zeo)::RecET-SacB to create a clpP1P2 merodiploid. The pEXPR(kan) plasmid was generated using multi-site gateway to insert the tetracycline-off repressor (RT38, PCR amplified using primers: attB4-RT38F, attB1-RT38R), the tetracycline-off promoter (p750, PCR amplified using primers: attB1-p750F, attB2-p750R), and the DAS-modified clpP1P2 operon (P1P2DAS, 2 rounds of PCR used to append DAS tag using primers: attB2-clpP1P2DASF, clpP1P2DASR, attB3-DASR) into the multisite gateway entry vector, pEN23A(kan). The DAS modification, which appends the DAS tag to the 3'-end of clpP2, was inserted to facilitate inducible degradation of ClpP2 upon expression of SspB. The plasmid pNIT(zeo)::RecET-SacB, which contains the machinery for mycobacterial recombineering, was a gift from Dirk Schnappinger. Next, hygromycin-containing knockout-out fragments were generated using stitch PCR to ligate 500 bp upstream (primers: 5flankF, 5flankR) and downstream (primers: 3flankF, 3flankR) of the clpP1P2 operon to a hygromycin resistance marker (primers: hygkoF, hygkoR). This linear PCR fragment was transformed into the above Mtb strain that was made competent for recombineering. This was done by addition of 1 mM isovaleronitrile (IVN, Sigma Aldrich) to a culture at OD_{600} 0.8. IVN addition induced expression of the recombineering machinery on pNit(zeo)::RecET-SacB. After 8h, 10 mL of 2M glycine was added, and the culture was grown overnight to yield recombineering-competent Mtb. Electroporation was performed as previously described[31]. Positive clones were plated on 7H10 agar containing 10% sucrose to counterselect against the recombinase plasmid, and scored for growth on zeocin-containing agar to confirm the loss of pNit::RecET. Deletion of the endogenous clpP1P2 operon was confirmed by PCR screening using primers that annealed outside of the region deleted and to the hygromycin resistance marker (primers: clpPkoF, hygR). In a final step, the tweety site integrating plasmid, pGMCtKq22(zeo)::TSC10M1-pUV15-sspB, was transformed into the strain. This plasmid, a gift from Dirk Schnappinger, inducibly expressed SspB and enabled degradation of the DAS-modified ClpP2 protein,

To construct GFP whole protein fusions, tetracycline-inducible, Gateway destination vectors were generated that enabled either N- or C-terminal fusion of an Mtb open reading frame to GFP. For N-terminal tagging, stitch PCR was used to append GFP, with a ribosome binding site upstream of the translational start site, (primers: RMR257, RMR258) upstream of an attR ccdB-Cm^R cassette (primers: RMR259, RMR260). For C-terminal tagging, stich PCR was used to append GFP (primers: RMR263, RMR264) downstream of an attR ccdB-CM^R cassette (primers: RMR261, RMR262). Both products were digested with PacI and EcoRV and then ligated into the tetracycline-inducible episomal pTetOR(zeo) plasmid, a gift from Mike Chao, to yield the final destination vectors. For N-terminal GFP tagging to carD, whiB1, rpL28, and dnaA, genes were amplified using primers containing attB sites (primers: RMR271,RMR272 [carD]; RMR275, RMR276 [whiB1]; RMR279, RMR280 [rpL28]; RMR283, RMR284 [dnaA]). The only difference for C-terminal tagging of Mtb genes was that the 5' end primer contained a ribosome-binding site between the attB1 sequence and ATG start of the gene (primers: RMR273, RMR274 [carD], RMR277, RMR278 [whiB1], RMR281, RMR282 [rpL28], RMR285, RMR286 [dnaA]).

The constitutively expressing GFP fusions, bearing the truncated C-termini of WhiB1 or CarD were generated using primers to amplify GFP, in which the reverse complement primer contained 3' end bases of carD or whiB1 (primers: RMR248, RMR249, RMR294, RMR295, RMR296, RMR298). These modified gfp PCR products were then digested with PacI and ClaI and ligated into the episomal plasmid pMV762(zeo), a gift from Meera Unnikrishnan, in which inserted products were regulated by the constitutive groEL promoter.

The luciferase reporter to monitor whiB1 promoter activity was generated using stich PCR to ligate the 500 bp upstream of whiB1 (primers: RMR313, RMR310) to the firefly luciferase gene (primers: RMR311, RMR312). This fused construct was then digested with XbaI and HindII and ligated into pGH100A(hyg), a promoter-less plasmid that integrates into the chromosome at the Giles integration site.

Immunoblotting

Total protein lysates were prepared from equivalent cell numbers (determined by OD_{600} values) using bead beating and immunoblot analysis was performed. For figure 3.1C, Mtb cultures were grown for the specified times from a starting OD 0.05, in the presence of 1.5 μg/mL ATc. For figure 3.2A, Mtb cultures were grown for 48 hours from a starting OD 0.02, in the presence or absence of 1.5 μg/mL ATc. In Figure 3.3, Msm cultures were grown to from starting OD 0.05 for 3-18 hours in fresh medium with or without ATc. The primary polyclonal antibody anti-ClpP2 (gift from Olga Kandror and Alfred Goldberg) was used at a 1:10,000 dilution in 5% milk powder in TBS-T, and the primary monoclonal anti-myc (Sigma Aldrich) was used as specified by manufacturer. Detection was performed using secondary HRP conjugated antibodies (Pierce), and SuperSignal West Femto Chemiluminescent Substrate (ThermoScientific) according the manufacturer's protocol. In all cases, blots were reprobed with anti-RpoB (MyBiosource), or initial protein gels were stained with SimplyBlue SafeStain (Life Technologies), to ensure equivalent loading of samples.

Fluorescence and Luminescence Measurements

In order to measure the abundance of GFP fusion proteins, cultures within one experiment were normalized based on OD_{600} values, spun down to remove media, and resuspended in 100 uL of PBS in a clear bottom 96 well plate. For luminescence, cultures were normalized based on OD_{600} values, and 150 uL of culture was used for measuring luciferase activity. 50 uL of Cell Culture 5X Lysis Reagent (Promega) was added to cultures, and samples were agitated for 10min on an orbital shaker, at room temperature. Next, 75 uL of Luciferase Assay Substrate (Promega) was added to each sample and directly taken for measurement. Fluorescence was measured at 485/538 nm, and luminescence was measured at an exposure time of 10 milliseconds, by the Fluroskan Ascent FL plate reader (ThermoScientific). Results represent the median +/- standard deviation of biological replicates.

Quantitative PCR

In Mtb and Msm, RNA was generated from equivalent of 20 mLs of cells at OD_{600} 0.5. Cultures were spun down, and subjected to bead beating (3X 45 sec each, 5 min on ice between pulses) after resuspension in TRIzol. After chloroform phase separation, genetic material was precipitated with isopropanol, resuspended in dH_2O, and RNA was purified using RNeasy Mini Kit (Qiagen). To ensure no contamination from genomic DNA, purified RNA was subjected to an additional round of DNase digestion using the TURBO DNA-free Kit (Invitrogen). cDNA was created from isolated equal concentrations of RNA with the SuperScript III First Strand Synthesis System (Invitrogen). Quantitative PCR was performed with the SYBR FAST qPCR kit (Kapa Biosystems) using the Applied Biosystems 7500 Fast Real-Time PCR System. Primers for each gene were designed using Primer3 (http://frodo.wi.mit.edu/), to have a melting temp of 60°C and a product length of 100-120 bp. All experiments were done using biological replicates, and representative experiments are depicted +/- standard error of mean of technical replicates.

Sample preparation, protein digestion, and peptide TMT labeling

For Mtb proteomics, P750-clpP1P2DAS was diluted to a starting OD_{600} 0.02 in 900 mLs of 7H9 media. This culture was split into six 150 mL cultures, and 1.5 μg/mL ATc was added to three batches, while three were left to grow with induction. After 48 hours, cultures were spun down (10min, 4000rpm, 4°C) and washed 3X with PBS. Cultures were then resuspended in 1 mL Urea Lysis Buffer (8M urea in 50 mM Tris pH 8.2, 75 mM NaCl, 50 mM NaF, 50 mM β-glycerophosphate, 1mM Na orthovanadate, Roche Complete EDTA-free Protease Inhibitor Cocktail tablets), and subjected to bead beating (3X 45 sec each, 5 min on ice between pulses). Cell lysates were spun down (10 min, 13,000rpm, 4°C), and samples were reduced with DTT (final concentration of 5 mM) for 30 min at 37°C. Samples were then alkylated with iodoacetamide (final concentration of 14 mM) for 30 min at room temperature in the dark. Adding an additional 5 mM DTT and incubating samples at room temperature in the dark for 15 min quenched excess

iodoacetamide. To remove samples from the BL3-level facility, proteins were precipitated with 20% trichloroacetic acid and incubated on ice for an hour. Proteins were pelleted by centrifugation (30 min, 13,000rpm, 4°C), and pellets were washed twice with acetone. Samples were resuspended in 8M urea, diluted, and protein amounts were quantified using a BCA assay (ThermoScientific).

As the proteomics screen in Msm was performed prior to the development of MS3-based proteomics for TMT analysis, we performed MS2-based quantitation of TMT peptide signals. For Msm proteomics, Msm/pTet(OR)::HIV2pr and clpP2_ID Msm were diluted to a starting OD_{600} 0.05 in 450 mLs. ATc (100 ng/mL) was added to each strain, and cultures were divided into 150 mL batches. Samples were harvested at 0h, 5h, and 11h post addition of ATc, spun down (10min, 4000rpm, 4°C), and washed 3X with PBS. Protein samples were prepared in a similar fashion as above, except no TCA precipitation was needed, and protein quantitation by the BCA assay was done prior to sample reduction and alkylation.

Proteins isolated from Mtb and Msm were digested overnight with Lys-C (Wako) in a 1:100 enzyme:protein ratio in 4M urea and 50 mM Tris-HCl (pH 8.2). Digests were acidified with formic acid to a pH of ~2-3, and subjected to C_{18} solid-phase extraction (Sep-Pak, Waters). Isobaric labeling of the peptides was accomplished with sixplex TMT reagents (Thermo Scientific). Reagents (0.8 mg) were dis- solved in 40 µl acetonitrile, and 10 µl of the solution was added to 250 µg of peptides dissolved in 100 µl of 50 mM HEPES (pH 8.5). After 1 h at room temperature (22 °C), the reaction was quenched by adding 8 µl of 5% hydroxylamine. Half of each of the labeled reactions was pooled into one vial, and combined fractions were subjected to C_{18} solid-phase extraction.

Sample fractionation by high pH reverse phase and strong cation exchange chromatography

Fractionation of Mtb TMT-labeled peptides was performed by high pH reverse phase liquid chromatography. Briefly, the sample was resuspended in high pH buffer A (10 mM ammonium formate pH 8.5, 5% acetonitrile), and separated over a 4.6 mm X C-18 analytical HPLC column (5 µm, 300 Å, Agilent). Separation involved a two-buffer (high pH buffers A and B) gradient from 0% to 20% high pH buffer B (10 mM ammonium formate pH 8.5, 96% acetonitrile) in 5 min at a flow rate of 0.8 ml min-1, followed by 20% to 45% high pH buffer B in 50 min at a flow rate of 0.8 ml min-1, followed by 45 to 100% high pH buffer B in 4 min at a flow rate of 0.8 ml min-1 using an Agilent 1100 quaternary pump outfitted with a degasser and a photodiode array detector (PDA) (Thermo Scientific). Samples were collected in 30-s increments into a 96-well plate and dried under vacuum. For even columns in the 96 well plate, wells A,C,E,G were combined into one fraction, while for odd columns, wells B,D,F,H were combined, resulting in 12 total fractions, which were dried under vacuum. Fractions were then redissolved with 1% formic acid, desalted by C_{18} solid phase extraction and re-dried under vacuum.

Fractionation of Msm TMT-labeled peptides was performed by SCX chromatography. Briefly, the sample was resuspended in SCX chromatography buffer A (7 mM KH2PO4 (pH 2.6) and 30% acetonitrile) and separated over a 4.6 mm × 200 mm polysulfoethyl A HPLC column (5 µm, 200 Å, PolyLC). Separation involved a two-buffer (SCX chromatography buffers A and B) gradient from 0 to 50% SCX chromatography buffer B (7 mM KH2PO4, 350 mM KCl (pH 2.6) and 30% acetonitrile) in 47 min at a flow rate of 0.9 ml min−1, followed by 50 to 100% SCX chromatography buffer A to buffer B in 4.5 min using an Agilent 1100 quaternary pump outfitted with a degasser and a photodiode array detector (PDA) (Thermo Scientific). Samples were collected in 30-s increments into a 96-well plate and dried under vacuum. Fractions were then redissolved with 1% formic acid and, based on the intensity from the SCX chromatographic UV-light trace, combined into a total of 20 samples of similar peptide amount, which were desalted by C_{18} solid phase extraction and dried under vacuum.

Liquid chromatography electrospray ionization tandem mass spectrometry

All proteomic methodology was adapted from a methods paper published on comparing MS2 to MS3 based quantitative proteomics using TMT isobaric labeling[23]. Briefly, all LC-MS2(MS3) experiments were performed on an LTQ Orbitrap Velos (Thermo Fischer Scientific) equipped with a Famos autosampler (LC Packings) and an Agilent 1100 binary high-pressure liquid chromatography (HPLC) pump (Agilent Technologies). Peptides were separated on a 100 µm inner diameter microcapillary column packed first with approximately 0.5 cm of Magic C4 resin (5 µm, 100 Å, Michrom Bioresources) followed by 20 cm of Maccel C18AQ resin (3 µm, 200 Å, Nest Group). Separation was achieved by applying a 9–32% acetonitrile gradient in 0.125% formic acid over 150 min at ~300 nl min−1. Electrospray ionization was enabled through applying a voltage of 1.8 kV through a polyetheretherketone (PEEK) junction at the inlet of the microcapillary column.

The LTQ Orbitrap Velos was operated in data-dependent mode for both MS2 and MS3 methods. For the MS2 method, the survey scan was performed in the Orbitrap in the range of 300–1,500 *m/z* at a resolution of 3 × 104, followed by the selection of the ten most intense ions (top 10) for HCD-MS2 fragmentation using a precursor isolation width window of 2 *m/z*. The automatic gain control (AGC) settings were 3 × 106 ions and 2.5 × 105 ions for survey and MS2 scans, respectively. Ions were selected for MS2 when their intensity reached a threshold of 500 counts and an isotopic envelope was assigned. Maximum ion accumulation times were set to 1,000 ms for survey MS scans and to 250 ms for MS2 scans. The normalized collision energy for HCD-MS2 experiments was set to 45% at a 30-ms activation time. Singly charged ion species and ions for which a charge state could not be determined were not subjected to MS2. Ions within a 10 p.p.m. *m/z* window around ions selected for MS2 were excluded from further selection for fragmentation for 120 s.

The survey MS scan settings were identical for the MS3 method, where the ten most intense ions were first isolated for ion trap CID-MS2 at a precursor ion isolation width of 2 *m/z*, using an AGC

setting of 2 × 103, a maximum ion accumulation time of 150 ms and wide band activation. Directly after each MS2 experiment, the most intense fragment ion in an *m/z* range between 110–160% of the precursor *m/z* was selected for HCD-MS3. The fragment-ion isolation width was set to 4 *m/z*, the MS3 AGC was 20 × 104 and the maximum ion time 250 ms. We chose an isolation width of 4 *m/z* to avoid space charging in preliminary experiments with very high AGC settings for the MS3 scan; but with the AGC set- tings used in the experiments described here, setting 2 or 4 *m/z* as the MS3 isolation width had a negligible effect on the results (data not shown). Normalized collision energy was set to 35% and 60% at an activation time of 20 ms and 50 ms for MS2 and MS3 scans, respectively. It is important to note that charge state screening has to be disabled to allow fragment ions to be selected for MS3; this setting nevertheless allows for charge state-based exclusion of singly charged ions and ions for which no charge state was determined from MS2.

Data processing: MS2 spectra assignment, data filtering and quantitative data analysis

A suite of in-house–developed software tools was used to convert mass spectrometric data from the RAW file to the mzXML format, as well as to correct erroneous assignments of peptide ion charge state and monoisotopic *m/z* (ref. 13). We modified the ReAdW.exe to include ion accumulation time in the output during conver- sion to the mzXML file format (http://sashimi.svn.sourceforge. net/viewvc/sashimi/). Assignment of MS2 spectra was performed using the Sequest algorithm by searching the data against a protein sequence database containing all known translated proteins from either Mtb or Msm, and known contaminants such as porcine trypsin, and human keratin[32]. This forward (target) database component was followed by a decoy component including all listed protein sequences in reversed order. Searches were performed using a 50 p.p.m. precursor ion tolerance, where both peptide termini were required to be consistent with Lys-C specificity, while allowing up to two missed cleavages. Sixplex TMT tags on lysine residues and peptide N termini (+ 229.162932 Da) and carbamido- methylation of cysteine residues (+57.02146 Da) were set as static modifications, oxidation of methionine

residues (+ 15.99492 Da) as a variable modification. An MS2 spectral assignment false discovery rate of less than 1% was achieved by applying the target- decoy database search strategy[33]. Filtering was performed using a linear discrimination analysis method to create one combined filter parameter from the following peptide ion and MS2 spectra properties: Sequest parameters XCorr and ΔCn, peptide ion mass accuracy and charge state, predicted low pH (2.7) in-solution charge of peptide and peptide length. Linear discrimination scores were used to assign probabilities to each MS2 spectrum for being assigned correctly and these probabilities were used to filter the dataset with an MS2 spectra assignment false discovery rate of less than 1% to obtain a protein identification false discovery rate of less than 1.5%[34].

For quantification, a 0.06 *m/z* window centered on the theoretical *m/z* value of each reporter ion was monitored for ions, and the intensity of the signal closest to the theoretical *m/z* value was recorded. Reporter ion intensities were denormalized by multiplication with the ion accumulation time for each MS2 or MS3 spectrum and adjusted based on the overlap of isotopic envelopes of all reporter ions. Intensity distributions of isotopic envelopes were as provided by the manufacturer. The total signal intensity across all peptides quantified were summed for each TMT channel, and all intensity values were normalized to account for potentially uneven TMT labeling. For MS3-based Mtb proteomics, the intensities for all peptides of a given protein were summed to derive an overall protein abundance value for each TMT signal. For MS2-based Msm proteomics, the overall protein abundance was calculated by taking the median intensity value of all peptides of that protein within each TMT channel. Hierarchical clustering using Pearsons correlational analysis was conducting using MultiExperiment Viewer (TM4 Microarray Software Suite)[19].

WORKS CITED

1. Glickman, M. H. & Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* **82**, 373–428 (2002).

2. Goldberg, A. L. Functions of the proteasome: from protein degradation and immune surveillance to cancer therapy. *Biochem. Soc. Trans.* **35**, 12–17 (2007).

3. Baker, T. A. & Sauer, R. T. ATP-dependent proteases of bacteria: recognition logic and operating principles. *Trends Biochem Sci* **31**, 647–653 (2006).

4. Pallen, M. J. & Wren, B. W. The HtrA family of serine proteases. *Molecular Microbiology* **26**, 209–221 (1997).

5. Gerth, U. *et al.* Clp-dependent proteolysis down-regulates central metabolic pathways in glucose-starved Bacillus subtilis. *J Bacteriol* **190**, 321–331 (2008).

6. Hong, S.-J., Lessner, F. H., Mahen, E. M. & Keiler, K. C. Proteomic identification of tmRNA substrates. *Proc Natl Acad Sci USA* **104**, 17128–17133 (2007).

7. World Health Organization Global tuberculosis control: WHO report 2011. Geneva, Switzerland: World Health Organization; 2011. (2011).

8. Damerau, K. & St John, A. C. Role of Clp protease subunits in degradation of carbon starvation proteins in Escherichia coli. *J Bacteriol* **175**, 53–63 (1993).

9. Gerth, U., Kruger, E., Derré, I., Msadek, T. & Hecker, M. Stress induction of the Bacillus subtilis clpP gene encoding a homologue of the proteolytic component of the Clp protease and the involvement of ClpP and ClpX in stress tolerance. *Molecular Microbiology* **28**, 787–802 (1998).

10. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology* **48**, 77–84 (2003).

11. Raju, R. M. *et al.* Mycobacterium tuberculosis ClpP1 and ClpP2 Function Together in Protein Degradation and Are Required for Viability in vitro and During Infection. *PLoS Pathog* **8**, e1002511 (2012).

12. Akopian, T. *et al.* The active ClpP protease from M. tuberculosis is a complex composed of a heptameric ClpP1 and a ClpP2 ring. *EMBO J* 1–13 (2012).doi:10.1038/emboj.2012.5

13. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**, 27–43 (1999).

14. Ribeiro-Guimarães, M. L. & Pessolani, M. C. V. Comparative genomics of mycobacterial proteases. *Microb Pathog* **43**, 173–178 (2007).

15. Klotzsche, M., Ehrt, S. & Schnappinger, D. Improved tetracycline repressors for gene silencing in mycobacteria. *Nucleic Acids Res* **37**, 1778–1788 (2009).

16. Kim, J.-H. *et al.* Protein inactivation in mycobacteria by controlled proteolysis and its application to deplete the beta subunit of RNA polymerase. *Nucleic Acids Res* (2010).doi:10.1093/nar/gkq1149

17. van Kessel, J. C. & Hatfull, G. F. Mycobacterial recombineering. *Methods Mol Biol* **435**, 203–215 (2008).

18. Dayon, L. *et al.* Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Anal Chem* **80**, 2921–2931 (2008).

19. Saeed, A. I. *et al.* TM4 microarray software suite. *Meth. Enzymol.* **411**, 134–193 (2006).

20. Barik, S., Sureka, K., Mukherjee, P., Basu, J. & Kundu, M. RseA, the SigE specific anti-sigma factor of Mycobacterium tuberculosis, is inactivated by phosphorylation-dependent ClpC1P2 proteolysis. *Molecular Microbiology* (2009).doi:10.1111/j.1365-2958.2009.07008.x

21. Smith, L. J. *et al.* Mycobacterium tuberculosis WhiB1 is an essential DNA-binding protein with a nitric oxide-sensitive iron-sulfur cluster. *Biochem J* **432**, 417–427 (2010).

22. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. & Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**, 671–683 (2003).

23. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* **8**, 937–940 (2011).

24. Singh, A. *et al.* Mycobacterium tuberculosis WhiB3 responds to O2 and nitric oxide via its [4Fe-4S] cluster and is essential for nutrient starvation survival. *Proc Natl Acad Sci USA* **104**, 11562–11567 (2007).

25. Garg, S. K., Suhail Alam, M., Soni, V., Radha Kishan, K. V. & Agrawal, P. Characterization of Mycobacterium tuberculosis WhiB1/Rv3219 as a protein disulfide reductase. *Protein Expression and Purification* **52**, 422–432 (2007).

26. Steyn, A. J. C. *et al.* Mycobacterium tuberculosis WhiB3 interacts with RpoV to affect host survival but is dispensable for in vivo growth. *Proc Natl Acad Sci USA* **99**, 3147–3152 (2002).

27. Reddy, T. B. K. *et al.* TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res* **37**, D499–D508 (2009).

28. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**, 1338–1347 (1998).

29. Stallings, C. L. *et al.* CarD Is an Essential Regulator of rRNA Transcription Required for Mycobacterium tuberculosis Persistence. *Cell* **138**, 146–159 (2009).

30. Jenal, U. & Fuchs, T. An essential protease involved in bacterial cell-cycle control. *EMBO J* **17**, 5658–5669 (1998).

31. van Kessel, J. C. & Hatfull, G. F. Recombineering in Mycobacterium tuberculosis. *Nat Methods* **4**, 147–152 (2006).

32. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **5**, 976–989 (1994).

33. Elias, J. & Gygi, S. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207–214 (2007).

34. Huttlin, E. L. *et al.* A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell* **143**, 1174–1189 (2010).

Chapter 4:

Targeting Clp protease for the development of new antibiotics effective against *Mycobacterium tuberculosis*

Section 4.1 – Chapter Overview and Attributions

Given its essentiality, and that genetic depletion of the enzyme had a rapid bactericidal effect, we believed Clp protease to be an attractive therapeutic target against *Mycobacterium tuberculosis* (Mtb). In this chapter, we partnered with collaborators who developed a series of β-lactone compounds that inhibit Clp protease in *Staphylococcus aureus* and *Listeria monocytogenes*. In these organisms, Clp is not required for in vitro growth, but is necessary for virulence, and so the β-lactones are currently being pursued as anti-virulence agents for Gram-positive pathogens. As Clp is essential for normal growth in mycobacteria, we wanted to evaluate their potential as traditional growth inhibitors in Mtb. One β-lactone, EZ120, had a superb killing profile, with an MIC of 1.6 μM and an MBC between 3-6 μM. Furthermore, we were able to show that EZ120 inhibited ClpP1P2 proteolysis in an in vitro purified protein assay, and that the compound interacted with the enzyme complex in intact cells. Together, this evidence validates the β-lactones as a potential class of antimycobacterials and Clp protease as a suitable target for more focused drug development.

Attributions. I drafted the manuscript that forms this chapter, which we aim to submit shortly as a brief communications to Nature Chemical Biology. The article will be a joint first-authored piece, with my close collaborator, Evelyn Zeiler, who is a graduate student in the laboratory of Professor Stephan Sieber at the Technical University Munich. The β-lactone compounds used in this study were synthesized and graciously provided by Evelyn (along with the detailed synthesis protocols below). I performed all MIC and MBC testing of the compounds in Mtb and Msm. I also conducted the in vitro and in vivo activity based protein profiling (ABPP) labeling experiments, and sent prepped samples to Evelyn who did the mass spectrometry workup and analysis. In vitro activity assays were carried out with tremendous help from Tatos Akopian, a senior scientist in the laboratory of Professor Alfred Goldberg. Finally, Evelyn performed the testing of β-lactone toxicity J774 mouse macrophages.

Section 4.2 – β-lactones kill *Mycobacterium tuberculosis* via inhibition of proteolysis by the essential Clp protease

ABSTRACT

Here, we show that β-lactones have potent activity against *Mycobacterium tuberculosis* (Mtb). These compounds inhibit proteolysis in vitro by the ClpP1P2 proteolytic core from Mtb, and interact with the protease in whole cells. Together, this data demonstrates that β-lactones exhibit a bactericidal effect in Mtb through inhibition of Clp protease, and suggests a unique mechanism of action compared to traditional protease suicide inhibitors.

MAIN ARTICLE

There is an urgent need to discover novel drugs with activity against *Mycobacterium tuberculosis* (Mtb), the causative agent of tuberculosis. Almost more troubling than the 1.3 million deaths annually[1], is the significant proportion of mortality now caused by multi- and extremely-drug resistant strains (MDR-TB, XDR-TB). In 2010 alone, there was an estimated MDR-TB prevalence of 650,000, accounting for 150,000 deaths[2]. To make matters worse, totally-drug resistant infections have been documented[3,4], and in the face of inadequate surveillance and ineffective drug regimens, these strains will surely proliferate. Chemotherapeutic options to deal with this challenge are limited; so much so that surgical intervention is now being employed, as was mandated in the pre-antibiotic era[5]. There are a few reports of drugs in clinical trials and lead candidates[6], but there is still a significant need to identify small molecules that exploit novel pathways and can serve as starting points for new antimycobacterial drugs.

In certain Gram-positive pathogens, such as *Staphylococcus aureus*, β-lactones have been found to abolish virulence through inhibition of Clp protease[7]. Though not essential for growth in defined

media, Clp is required for virulence factor production, presumably through proteolytic action on virulence-associated transcriptional repressors[8]. Beta-lactones had no MIC in *S. aureus*, but did phenocopy Clp depletion by abolishing secretion of several key virulence factors and severely attenuating virulence[9]. Furthermore, activity-based protein profiling (ABPP) using derivatized β-lactones demonstrated that the compounds interacted with the ClpP proteolytic subunit.

In bacteria, Clp is a compartmentalized protease found as a complex of proteolytic and ATPase subunits[10]. Canonically, the proteolytic core is a tetradecamer of ClpP subunits, each which possesses a serine-based catalytic triad. This core is governed by a hexameric ATPase complex that interacts with the apical surface of the tetradecamer and regulates substrate entry for proteolysis. In most organisms, Clp is dispensable for normal growth, but it has recently been found that the complex is required for normal growth of mycobacteria, and that genetic depletion of the protease had a rapid bactericidal effect[11].

To assess whether this divergent essentiality of Clp could be exploited to yield β-lactones as traditional growth inhibitors in mycobacteria, we first tested the ability of a panel of derivatives **(FIGURE 4.1)** to inhibit proteolysis by the enzyme in vitro. In *Mycobacterium tuberculosis*, the proteolytic tetradecamer has a unique structure, and is a heteromer of two distinct subunits ClpP1 and ClpP2. We have previously reported the establishment of an in vitro Mtb ClpP1P2 peptidase assay[12]. Briefly, a small, di-peptide activator (Z-Leu-Leu) is added to equimolar ratios of ClpP1 and ClpP2 and proteolysis is monitored by cleavage of a flurogenic peptide substrate with a C-terminal aminomethylcoumarin group (Z-Gly-Gly-Leu-AMC). Incubation of the ClpP1P2 complex with a small panel of β-lactones revealed that two compounds, P1 and EZ120, significantly inhibited peptide hydrolysis at a concentration of 200 μM **(FIGURE 4.2A)**. Inhibition was dose-dependent, and plotting enzyme activity against drug concentration, revealed IC_{50} values of 10 μM and 75 μM for P1 **(FIGURE 4.2B)** and EZ120 **(FIGURE 4.2C)**, respectively.

In order to determine how the β-lactones interacted with the protease, we used alkyne-derivitized probe forms (EZ120P and P1, which was already derivatized) of the active compounds to label the tetradecamer in vitro **(FIGURE 4.2D)**. The alkyne handle permitted us to append an azide-rhodamine tag to the β-lactone probes via click chemistry (CC) after compound incubation with the purified proteins[13]. In vitro testing revealed that the probe EZ120P inhibited proteolysis to a similar degree as the parent compound, EZ120. SDS PAGE analysis revealed that EZ120P and P1 interacted with both ClpP1 and ClpP2. Labeling in vitro revealed three surprising findings. First, as the probes labeled individual preparations of ClpP1 and ClpP2 alone, it appears that heterotetradecameric formation is not required for β-lactone interaction. Second, the addition of activator appears to compete with β-lactone suggesting that the two may access the same site within the protease. The in vivo significance of the di-peptide activator is still disputed, but the presence of a mycobacterial specific activator of protease complex formation was suggested by our in vitro findings[12] and the inability to reconstitute Mtb ClpP1P2 in *Escherichia coli* whole cells and extracts with ClpP1 and ClpP2 alone. Third, the β-lactones were still able to bind ClpP1 and ClpP2 when serine in the catalytic triad was inactivated by alanine substitution. This was entirely unexpected as we believed the β-lactones to be suicide inhibitors that acted by irreversibly reacting with the active site serine. MS/MS analysis of the labeled complex revealed that the β-lactones reacted with ClpP1 and ClpP2 at C58 and C105, respectively **(FIGURE 4.2E, 4.3)**. Surprisingly, the nucleophilic serine in both subunits was unmodified. We believe this to be a specific interaction as other cysteines and nucleophilic residues in the ClpP1 and ClpP2 were not modified, and, furthermore, β-lactone incubation did have an effect on proteolysis suggesting that this modification had a functional consequence.

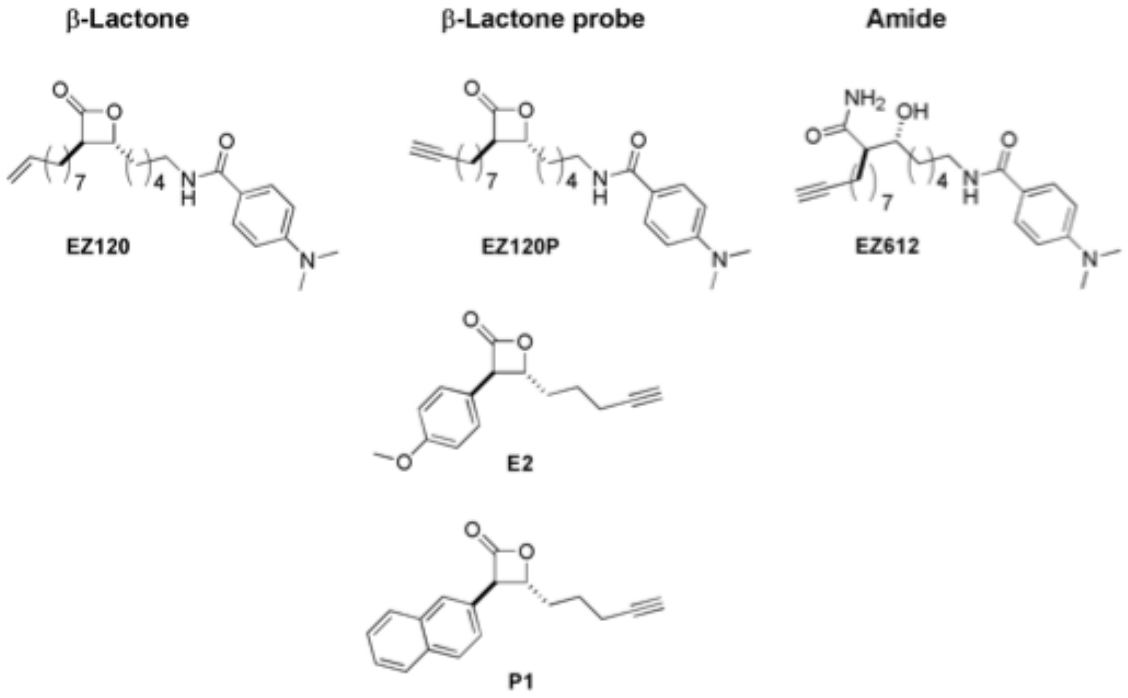


Figure 4.1 Structures of β-lactones used in this study.

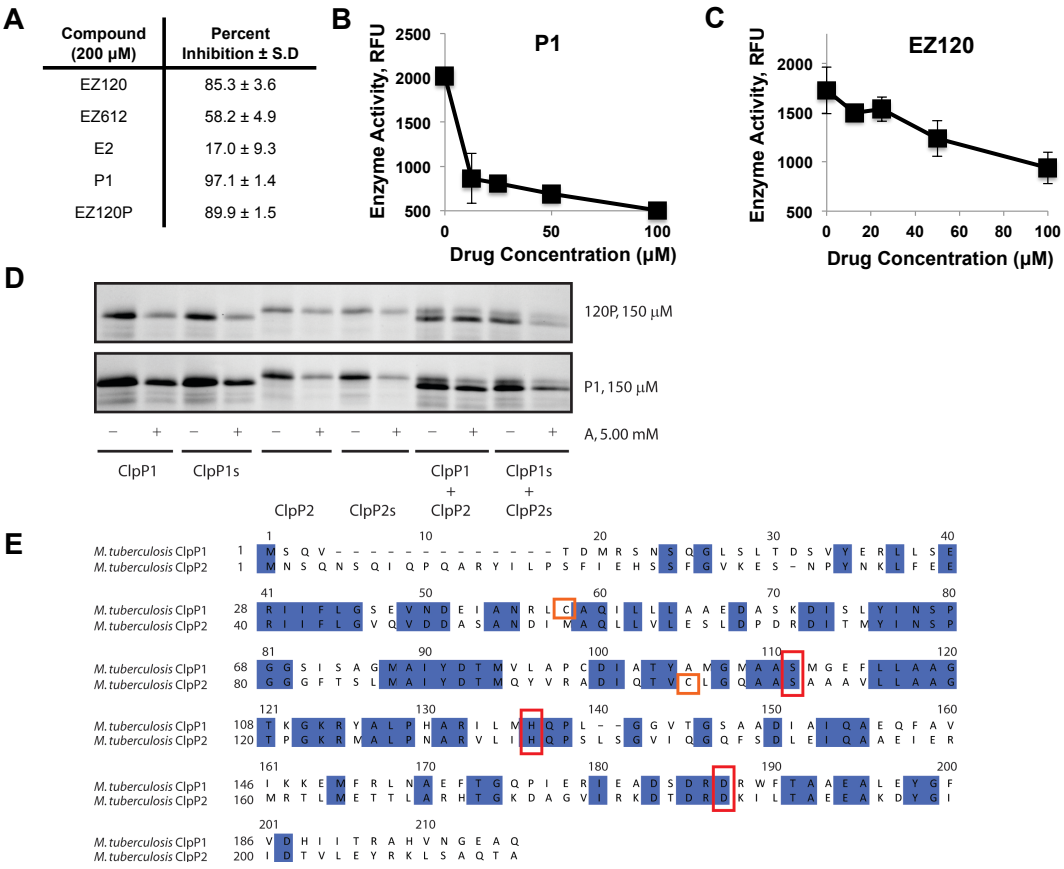


Figure 4.2 β-lactones react with purified Mtb ClpP1 and ClpP2,and inhibit proteolysis in vitro

(A) Inhibition of in vitro ClpP1P2 proteolytic activity on a fluorogenic peptide substrate was measured for a panel of β-lactones, at a concentration of 200 μM. Inhibitory potential of each β-lactone is reported as a percent inhibition +/- standard deviation, which is the average loss of enzyme activity over triplicate runs in the presence of compound divided by the average activity in the absence of compound. (B) Inhibitory potential for β-lactone P1 was assessed by determining ClpP1P2 activity on a fluorogenic peptide substrate (enzyme activity, RFU) as a function of drug concentration (μM). (C) Inhibitory potential for β-lactone EZ120, determined in the same manner as the compound P1. For (B) and (C), data are plotted as the average of triplicate runs +/- standard deviation. (D) Fluorescent SDS PAGE analysis of purified preparations of ClpP1, ClpP1S98A (ClpP1S), ClpP2, ClpP2S110A (ClpP2S), ClpP1P2, and ClpP1SP2S in the presence or absence of activator (denoted A), after incubation with EZ120P or P1 and click chemistry with a rhodamine-azide reagent. (E) ClpP1 and ClpP2 residues modified by β-lactones were identified by MS/MS analysis of purified proteins incubated with the compound P1 (150 μM). Modified sites that were identified are boxed in orange, while the residues that comprise the catalytic triad are boxed in red.

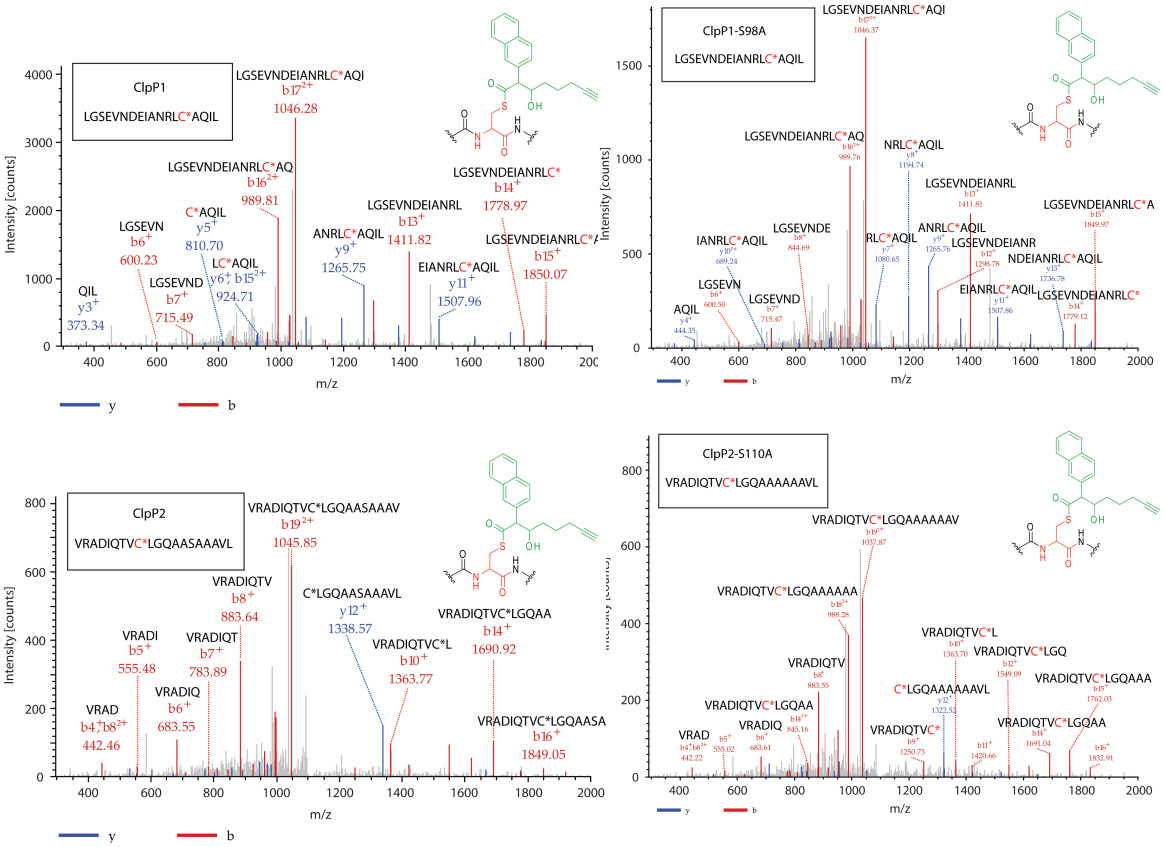


Figure 4.3 MS/MS analysis of purified protease subunits incubated with compound P1 in vitro.

MS2 spectra depicting the b and y ions for β-lactone modified peptides of ClpP1 (top left), ClpP1S98A (top right), ClpP2 (bottom left), and ClpP2S110A (bottom right). Spectra reveal that the β-lactone reactive residues are C54 and C105 in ClpP1 and ClpP2, respectively.

The inability of small molecules that exhibit promising activity in vitro to access the intracellular bacterial environment has largely thwarted drug discovery for novel antibiotics[14]. The permeability conundrum is especially pertinent for mycobacteria, which is largely resistant to many commonly used antibiotics due to a thick waxy-coat cell wall comprised of covalently linked peptidoglycan, arabinogalactan, and long chain fatty acids known as mycolic acids[15]. Having demonstrated an effect on ClpP1P2 proteolysis in vitro, we asked whether the β-lactones could penetrate mycobacterial defenses and access the essential protease in the bacterial cytoplasm. Minimum inhibitory concentration (MIC) determination of the β-lactones revealed that EZ120 effectively inhibited growth of mycobacteria with an MIC of 1.5 μM in Mtb and 50 μM *Mycobacterium smegmatis* (Msm), a fast-growing, non-pathogenic model organism, closely related to Mtb **(FIGURE 4.4A)**. The most active β-lactone in vitro, P1, did not affect the growth of mycobacteria, nor did the open-ring amide form of EZ120, EZ612.

Genetic depletion of the ClpP1P2 protease has a bactericidal effect in mycobacteria[11]. To determine if EZ120 phenocopied genetic perturbation and was also bactericidal in Mtb, we tested the compound's ability to kill actively growing Mtb. Varying concentrations of EZ120 were incubated with cultures at an initial inoculum of $4X10^5$ CFU/mL, and after three days the number of remaining viable bacteria in each culture was determined by plating for CFU/mL. **FIGURE 4.4B** reveals that EZ120 was potently bactericidal with nearly two logs of killing seen at concentrations as low as 3.125 μM, and no retrievable viable bacteria at 25 μM within three days. As a control, cultures exposed to varying concentrations of EZ612 revealed no effect on growth. The lack of an in vivo effect of EZ612 confirmed that the lactone moiety was critical for the activity of these compounds and further affirmed that the mechanism of action was related to irreversible interaction with nucleophilic amino acids.

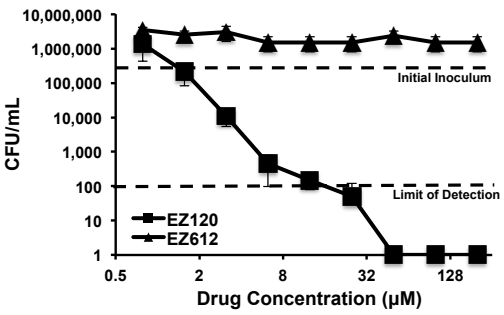
While the core structure of the compound appears to be important for killing in Mtb, multiple β-lactones have also been reported as inhibitors of the eukaryotic proteasome[16]. Significant

inhibition of the eukaryotic proteasome and the resulting toxicity would undoubtedly limit the therapeutic potential of the β-lactones. MTT testing of EZ120 in J774 macrophages, revealed an IC_{50} of 325 μM for metabolic inhibition **(FIGURE 4.4C)**, nearly 200 fold higher than the MIC of the compound in Mtb. Through robust determination of the therapeutic window of EZ120 will require further testing, this initial difference validates the pursuit of EZ120 as an initial hit for further development of a novel antimycobacterial compound.

A

Compound	MIC Mtb (μM)	MIC Msm (μM)
EZ120	1.6	50
EZ612	>200	>200
E2	>200	>200
P1	100	>200
EZ120P	1.6	50

B



C

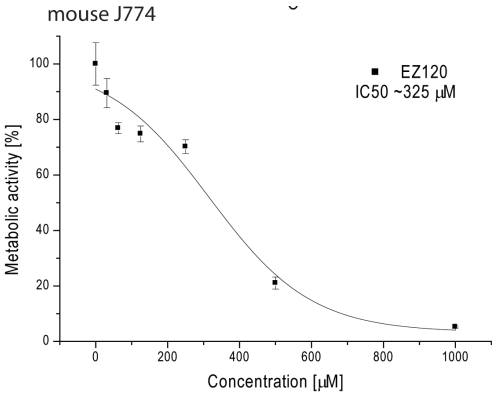
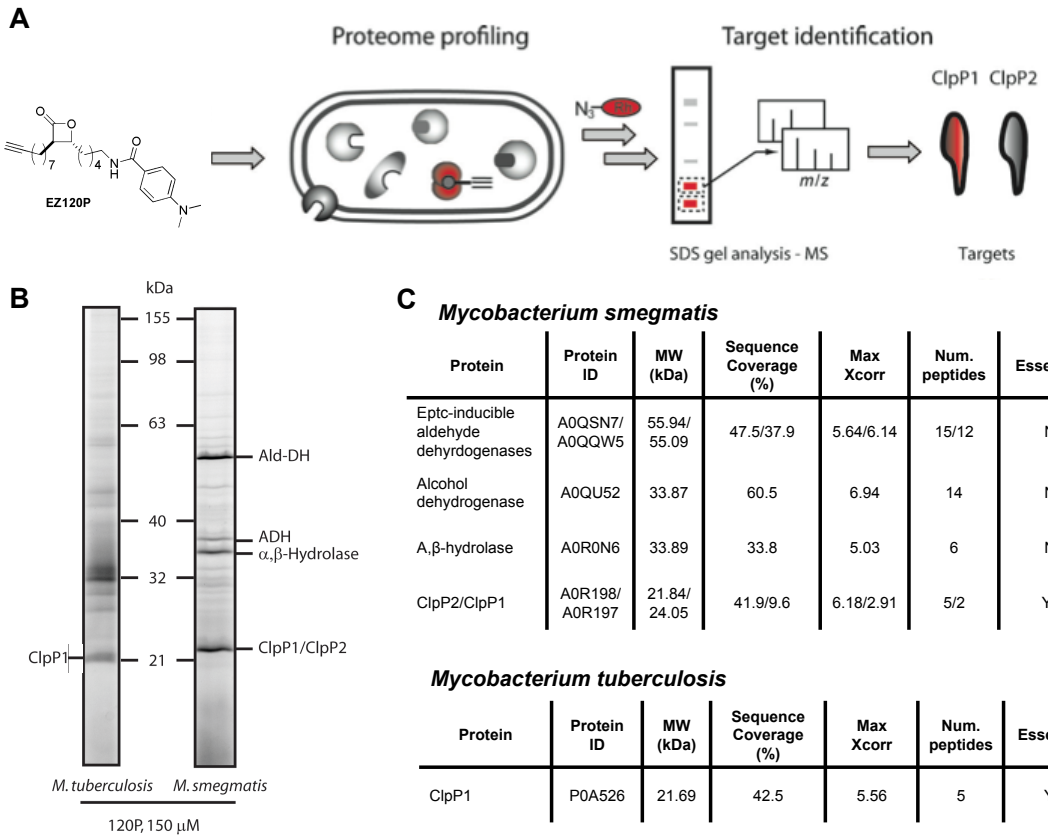


Figure 4.4 β-lactones are bactericidal in Mtb, and demonstrate a good therapeutic window.

(A) Minimum inhibitory concentrations (MICs) were determined for a panel of β-lactones in Mtb and Msm, using the colorimetric resazurin microtiter assay. (B) β-lactones EZ120 and EZ612 were assessed for their ability to kill actively growing Mtb. A mid-log phase culture of Mtb was diluted to an inoculum of $4X10^5$ CFU/mL (top dotted line) and incubated with a two-fold serial dilution of each lactone. After 3d, cultures were plated to determine viable bacteria, represented by CFU/mL. Data points are the average growth among biological replicates +/- standard deviation. (C) Cytotoxicity of EZ120 was measured in mouse macrophages (J774), using the MTT assay. Data points are represented as the average metabolic activity of biological triplicates +/- standard deviation.

Though EZ120 inhibited ClpP1P2 proteolysis in vitro, this was not conclusive evidence that the β-lactones exhibited the same effect in vivo. We were encouraged to pursue in vivo determination of the target based on the finding that EZ120P had a similar MIC in Mtb as the parent compound EZ120. Therefore, we turned to in situ ABPP experiments to identify the proteins that the active β-lactones interacted with in intact cells **(FIGURE 4.5A)**. Incubation of Mtb and Msm with EZ120P followed by cell lysis and click chemistry to append a rhodamine-bioitin-azide to β-lactone modified proteins showed that the β-lactones interacted with a variety of proteins in vivo **(FIGURE 4.5B)**. Streptavidin-based purification of tagged proteins, followed by MS/MS identification of major protein bands revealed that EZ120P bound several deyhdrogenases, an α/β-hydrolase and ClpP1P2 in Msm **(FIGURE 4.5C)**. Interestingly, the only essential proteins, as determined by genome wide transposon mutagenesis (Dragset et al, unpublished data), that we identified through in situ ABPP in Msm were ClpP1 and ClpP2. As one would expect that growth inhibition results from the inhibition of essential gene products, the presence of multiple β-lactone non-essential interacting proteins could be largely ignored. Further affirmation of Clp as the target of β-lactones resulted from identifying that the only overlapping band between Mtb and Msm, which are both sensitive to EZ120, was ClpP1.



C

Mycobacterium smegmatis

Protein	Protein ID	MW (kDa)	Sequence Coverage (%)	Max Xcorr	Num. peptides	Essential?
Eptc-inducible aldehyde dehyrdogenases	A0QSN7/ A0QQW5	55.94/ 55.09	47.5/37.9	5.64/6.14	15/12	No
Alcohol dehydrogenase	A0QU52	33.87	60.5	6.94	14	No
A,β-hydrolase	A0R0N6	33.89	33.8	5.03	6	No
ClpP2/ClpP1	A0R198/ A0R197	21.84/ 24.05	41.9/9.6	6.18/2.91	5/2	Yes

Mycobacterium tuberculosis

Protein	Protein ID	MW (kDa)	Sequence Coverage (%)	Max Xcorr	Num. peptides	Essential?
ClpP1	P0A526	21.69	42.5	5.56	5	Yes

Figure 4.5 Activity based protein profiling (ABPP) in situ demonstrates that EZ120P reacts with the Clp proteolytic core in Msm and Mtb.
(A) Schematic outline of ABPP experiments conducted to determine EZ120P interacting proteins in situ in Mtb and Msm. (B) Fluorescent SDS-PAGE analysis of the Mtb (left) and Msm (right) proteomes after in situ labeling with EZ120P, cell lysis, and click chemistry with a rhodamine-bioitin-azide reagent. (C) After click chemistry, samples were purified by streptavidin pull-down, and streptavidin-bound proteins were identified by MS/MS analysis. Major proteins from each band are denoted in (B), and details regarding MS/MS identification as well as essentiality of β-lactone interacting proteins are given in the table.

It is puzzling that Mtb in situ ABPP did not reveal an interaction with both ClpP1 and ClpP2. Even in Msm, EZ120P appeared to have a stronger affinity for ClpP1 compared to ClpP2. This in vivo preference may be reflected by our in vitro findings in Figure 4.2D, where equal amounts of protein and compound reveal that the β-lactones more strongly label ClpP1. There are some key differences in the protein sequence of ClpP1 and ClpP2, and these may translate to structural differences that could explain the preference of the β-lactones for ClpP1.

Given that certain β-lactones have been found to inhibit the eukaryotic proteasome, it is plausible that EZ120 may covalent modify the active site of the Mtb proteasome. We believe this is not the case, as none of the major or minor bands in Mtb or Msm analyzed by MS/MS corresponded to the proteasome subunits. Furthermore, evidence has emerged that proteolysis by the proteasome is dispensable for cell survival[17], and so inhibitors of the Mtb proteasome would not likely have the bactericidal effect observed with EZ120.

Several significant questions remain regarding the targeting of the ClpP1P2 proteolytic core by the β-lactones. How does the non-canonical β-lactone interaction result in ClpP1P2 inhibition and bacterial death ultimately? Our finding of discrete bands upon in situ ABPP further confirmed the specificity of this non-canonical β-lactone interaction with ClpP1 and ClpP2. The locations of these interactions reveal that they may block access of proteins and peptides to the active site or crowd the active sites and prevent substrate access to the catalytic residues. However, the absence of any structural information on the ClpP1P2 complex is limiting, and in order to determine this definitively, crystal structures of the β-lactones in complex with the tetradecamer must be solved. Visualization of the β-lactones with the ClpP1P2 complex would also elucidate potential reasons for the differential affinity of the compounds for ClpP1 compared to ClpP2.

In summary, we believe that the combination of in vitro enzyme activity and in situ activity based profiling, used here for the first time in Mtb, reveal that β-lactones exhibit a potent, bactericidal

effect through inhibition of the ClpP1P2 protease. Work is still needed to elucidate the mechanism of the non-canonical interaction we have observed, but these findings (along with the lack of toxicity in eukaryotic cell lines) validate for the first time the β-lactones and the essential mycobacterial Clp protease as the ideal therapeutic agent and target for the development of severely-needed antibiotics effective against *Mycobacterium tuberculosis.*

MATERIALS AND METHODS

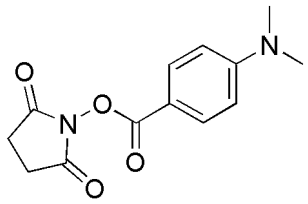
Synthesis of the EZ120 probe (EZ120P) and EZ612

All chemicals were of reagent grade or better and used without further purification. Chemicals and solvents were purchased from Sigma Aldrich and Acros Organics. For all reactions, only commercially available solvents of purissimum grade, dried over molecular sieve and stored under argon atmosphere were used. Solvents for chromatography and workup purposes were generally of technical grade and purified before use by distillation. Temperatures were measured externally. All experiments were performed under nitrogen or argon atmosphere. Column chromatography was performed on silica gel (Merck, 0.035–0.070 mm, mesh 60 Å). For TLC analysis, Merck TLC silica gel 60 F254 aluminium sheets were used. Preparative HPLC was performed on a Waters 2998 system with a XBridge Prep C18 column (5 µm, 30 x 150 mm) or a YMC-Triart C18 column (5 µm, 10 x 250 mm). ^{1}H NMR and ^{13}C NMR spectra were recorded on a Varian NMR-System 600 or 300 (600 MHz, 300 MHz), a Bruker AV 500 or AV 360 (500 MHz, 360 MHz) and referenced to the residual proton and carbon signal of the deuterated solvent, respectively. HRMS (ESI) was obtained by a LTQ-FT from Thermo Scientific.

The compounds **EZ120**[18], **U1**, **U1P**[19], **P1**, and **E2**[20] were synthesized as described before.

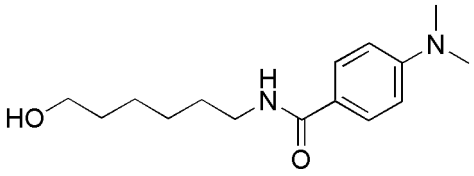
The β-Lactone **EZ120P** was prepared according to a method developed by Danheiser and Novick[21] and as applied before[9,18,20].

2,5-Dioxopyrrolidin-1-yl 4-(*N,N*-dimethylamino)benzoate (1) EZ100



4-(*N,N*-Dimethylamino)benzoic acid (332 mg, 2.00 mmol), *N*-hydroxysuccinimide (244 mg, 2.10 mmol) and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide, (EDAC) (402 mg, 2.1 mmol) were suspended in CH_2Cl_2 (10 mL). After 20 min the suspension became a clear violet solution. The reaction mixture was monitored by TLC (*iso*-hexane/ethyl acetate 1:1, 1% AcOH, R_f = 0.30) and the reaction was completed after 2.5 h. Then water (10 mL) was added and the aqueous layer was extracted with dichloromethane (2 x 10 mL). The combined organic layers were washed with brine (10 mL), dried over $MgSO_4$ and the solvent was removed under reduced pressure to yield 451 mg (86%) of **1** as a white solid. – Mp 181 °C. – ^{1}H NMR (300 MHz, *CDCl$_3$*) δ 7.99 (d, *J* = 9.1 Hz, 2 H, H*arom*), 6.67 (d, *J* = 9.2 Hz, 2 H, H*arom*), 3.09 (s, 6 H, $N(CH_3)_2$), 2.89 (s, 4 H, $(CH_2)_2$). – ^{13}C NMR (75 MHz, *CDCl$_3$*) δ 169.83, 161.92, 154.41, 132.55, 110.81, 110.65, 39.98, 25.70. – IR (film) ν = 1735 cm^{-1}, 1604, 1535, 1436, 1375, 1269, 1248, 1207, 1182, 1070, 1012, 979, 944, 825, 755. – HRMS (ESI) calcd. for $C_{13}H_{15}N_2O_4$ $[M+H]^+$ 263.1032, found 263.1028. The spectral data are identical to those published in literature[22].

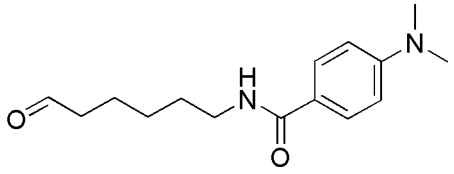
4-(*N,N*-Dimethylamino)-*N'*-(6-hydroxyhexyl)benzamide (2) EZ108



To a stirred solution of 2,5-dioxopyrrolidin-1-yl 4-(*N,N*-dimethylamino)benzoate (**1**) (850 mg, 3.24 mmol) in acetonitril/water 1:1.2 (11 mL) was added NEt_3 (2.72 mL, 19.7 mmol) and 6-amino-1-hexanol (380 mg, 3.24 mmol). The reaction was monitored by TLC (ethyl acetate, R_f = 0.24)

and was completed after 30 min. Then the mixture was extracted with ethyl acetate (3 x 40 mL), the combined organic layers were washed with brine (10 mL) and dried over $MgSO_4$. Purification by flash column chromatography on silica yielded 368 mg (43%) 4-(*N*,*N*-dimethylamino)-*N'*-(6-hydroxyhexyl)benzamide (**2**) as white solid. – Mp 101 °C. – ^{1}H NMR (360 MHz, *CDCl*$_3$) δ 7.69 (d, J = 8.8 Hz, 2 H, H*arom*), 6.66 (d, J = 8.9 Hz, 2 H, H*arom*), 6.25 (bs, 1 H, NH), 3.62 (t, J = 6.3 Hz, 2 H, NHC$\underline{H}_2$), 3.42 (m, 2 H, C$\underline{H}_2$OH), 3.00 (s, 6 H, N(CH_3)$_2$), 2.23 (bs, 1 H, OH), 1.67–1.51 (m, 4 H), 1.48–1.33 (m, 4 H). – ^{13}C NMR (91 MHz, *CDCl*$_3$) δ 167.55, 152.39, 128.29, 121.52, 111.09, 62.58, 45.82, 40.12, 39.62, 26.53, 8.60. – IR (film) ν = 3318 cm^{-1}, 2926, 2856, 2816, 1747, 1684, 1604, 1549, 1513, 1444, 1412, 1364, 1328, 1300, 1230, 1204, 1173, 1131, 1061, 946, 925, 872, 829, 807, 768. – HRMS (ESI) calcd. for $C_{15}H_{25}N_2O_2$ $[M+H]^+$ 265.1916, found 265.1907.

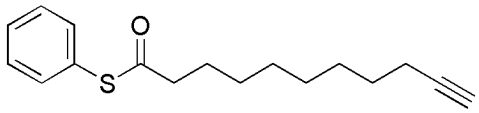
4-(*N*,*N*-Dimethylamino)-*N'*-(6-oxohexyl)benzamide (3) EZ112



To a solution of oxalyl chloride (60 µL, 0.72 mmol) in CH_2Cl_2 (2 mL) was added at –78 °C DMSO (102 µL, 1.44 mmol), 4-(*N*,*N*-dimethylamino)-*N'*-(6-hydroxyhexyl)benzamide (**2**) (159 mg, 0.60 mmol) in CH_2Cl_2 (3 mL) and NEt_3 (460 µL, 3.30 mmol). The reaction mixture was stirred at –78 °C for 5 h and was then warmed up to –5 °C. After work up and flash column chromatography on silica (*iso*-hexane/ethyl acetate 1:2) 103 mg (65%) of the product **3** was obtained as white solid, R_f = 0.27. – Mp 66 °C. – ^{1}H NMR (300 MHz, *CDCl*$_3$) δ 9.77 (t, J = 1.7 Hz, 1 H, CHO), 8.02–7.43 (m, 2 H, H*arom*), 6.71–6.65 (m, 2 H, H*arom*), 6.15 (bs, 1 H, NH), 3.45 (td, J = 7.0, 6.0 Hz, 2 H, NHC$\underline{H}_2$), 3.02 (s, 6 H, N(CH_3)$_2$), 2.46 (td, J = 7.2, 1.7 Hz, 2 H, C$\underline{H}_2$CHO), 1.75–1.57 (m, 4 H, (CH_2)$_2$), 1.49–1.35 (m, 2 H, CH_2). – ^{13}C NMR (75 MHz, *CDCl*$_3$) δ 202.55, 167.46, 152.36, 128.30, 121.53, 111.13, 43.74, 40.16, 39.50, 29.62, 26.45, 21.65. – IR (film) ν = 3335 cm^{-1}, 2922, 2859, 2818, 1717, 1685, 1607, 1542, 1513, 1444, 1411, 1363, 1328, 1299, 1229,

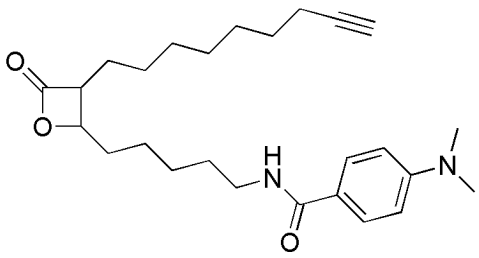
1203, 1173, 1134, 1064, 993, 946, 829, 768, 732. – HRMS (ESI) calcd. for $C_{15}H_{23}N_2O_2$ $[M+H]^+$ 263.1760, found 263.1755.

_S_-Phenyl-10-undecyne thioate (4) EZJS15



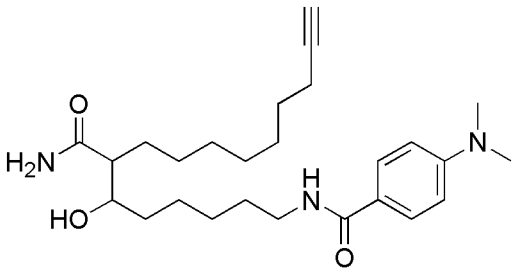
To a solution of 10-undecynoic acid (1.00 g, 5.49 mmol) in CH_2Cl_2 (28 mL) was added thiophenol (840 µL, 8.23 mmol), DCC (1.25 g, 6.06 mmol), and DMAP (67 mg, 0.55 mmol) and the reaction mixture was stirred for 16 h. The remaining precipitate was filtered off and the residue was washed with CH_2Cl_2 (50 mL). The solvent was removed under reduced pressure and purification by flash column chromatography on silica (hexane/ethyl acetate 25:1) afforded 1.15 g (76%) of the product **4** as a colourless oil, R_f = 0.50. – ^{1}H NMR (360 MHz, *CDCl$_3$*) δ 7.43 (s, 5H), 2.67 (t, *J* = 7.5 Hz, 1H), 2.21 (td, *J* = 7.0, 2.7 Hz, 2H), 1.96 (t, *J* = 2.6 Hz, 1H), 1.73 (p, *J* = 7.3 Hz, 2H), 1.62 – 1.48 (m, 2H), 1.48 – 1.28 (m, 5H). – ^{13}C NMR (91 MHz, *CDCl$_3$*) δ 197.50 , 134.47 , 129.28 , 129.14 , 84.69 , 68.13 , 43.69 , 29.10 , 28.88 , 28.86 , 28.65 , 28.42 , 25.55 , 18.39 . – IR (film) *v* = 2927 cm^{-1}, 2854, 1704, 1583, 1478, 1457, 1441, 1327, 1139, 1105, 1067, 1023, 962, 743, 705, 688. – HRMS (ESI) calcd. for $C_{17}H_{23}O_S$ $[M+H]^+$ 275.14696, found 275.14637. The spectral data are identical to those published in literature[19].

(3*R,4*R**)-4-(4-(*N,N*-Dimethylamino)benzamido-5-pentyl)-3-(non-8-ynyl)oxetan-2-one (EZ120P)**



A solution of diisopropylethylamine (122 µL, 0.863 mmol) in THF (7 mL) was cooled in an ice bath to 0 °C and under stirring *n*-butyllithium (540 µL, 0.863 mmol, 1.6 M in hexane) was gradually added via a syringe. The reaction mixture was stirred for 30 min at 0 °C. Then the ice bath was replaced with an acetone/dry ice bath and the mixture was cooled to –78 °C. A solution of *S*-phenyl-10-undecyne thioate (**4**) (189 mg, 0.690 mmol) in THF (4 mL) was injected dropwise over 15 min. After stirring the mixture at –78 °C for 2.5 h a solution of 4-(*N*,*N*-dimethylamino)-*N'*-(6-oxohexyl)benzamide (**3**) (181 mg, 0.690 mmol) in THF (3 mL) was added drop wise over 1 h via a syringe that was externally cooled by an aluminium funnel filled with dry ice. The reaction mixture was stirred for 2 h while gradually warmed up to 0°C. Then half saturated NH_4Cl solution (15 ml) was added and the mixture was diluted with ethyl acetate (50 ml). The aqueous layer was extracted with ethyl acetate (3 x 50 ml) and the combined organic layers were washed with K_2CO_3 solution (10%, 5 mL) and brine (5 ml) and dried over $MgSO_4$. The solvent was removed under reduced pressure to give 368 mg of crude product as yellow oil. Purification by flash column chromatography on silica (hexane/ethyl acetate 8:5) yielded 48 mg (16%) of the product as colourless oil, R_f = 0.14. – ^{1}H NMR (500 MHz, *CDCl$_3$*) δ 7.69 (d, J = 9.0 Hz, 2H), 6.70 (d, J = 9.0 Hz, 2H), 6.01 (t, J = 6.0 Hz, 1H), 4.24 (ddd, J = 7.6, 5.7, 4.0 Hz, 1H), 3.47 (qd, J = 7.0, 1.2 Hz, 2H), 3.19 (ddd, J = 8.6, 6.7, 4.0 Hz, 1H), 3.04 (s, 6H), 2.21 (td, J = 7.1, 2.6 Hz, 2H), 2.02 – 1.92 (m, 1H), 1.92 – 1.71 (m, 4H), 1.71 – 1.30 (m, 16H). – ^{13}C NMR (91 MHz, $CDCl_3$) δ 171.48, 167.43, 152.41, 128.25, 121.40, 111.08, 84.62, 77.96, 68.18, 56.18, 40.14, 39.60, 34.36, 29.74, 29.14, 28.76, 28.55, 28.36, 27.83, 26.89, 26.59, 24.87, 18.35. – IR (film) v = 2931 cm^{-1}, 2857, 1812, 1608, 1540, 1511, 1445, 1363, 1297, 1204, 1173, 1126, 1065, 947, 874, 830, 768, 743. – HRMS (ESI) calcd. for $C_{26}H_{39}N_2O_3$ $[M+H]^+$ 427.2961, found 427.2952.

(6*R,7*R**)-*N*-(7-Carbamoyl-6-hydroxy-hexadec–15–ynyl)–4–dimethylamino-benzamide (EZ612)**



A solution of (3*R**,4*R**)-4-(4-(*N,N*-Dimethylamino)benzamido – 5 - pentyl) – 3 - (non-8-ynyl)oxetan-2-one (**EZ120P**) (11 mg, 0.025 mmol) in 7 N NH_3/MeOH (1 mL) was stirred at rt for 1.5 h. The solvent was removed under reduced pressure and purification by flash column chromatography on silica (Ethyl acetate) yielded 6.2 mg (54%) of a white solid, R_f = 0.18. – Mp 93 °C. – ^{1}H NMR (360 MHz, *CDCl$_3$*) *δ* 7.67 (d, *J* = 8.9 Hz, 2H, H*arom*), 6.68 (d, *J* = 8.9 Hz, 2H, H*arom*), 6.62 (s, 1H, NH_2CO), 6.09 (t, *J* = 6.0 Hz, 1H, NHCO), 5.42 (s, 1H, NH_2CO), 3.68 – 3.41 (m, 5H, C<u>H</u>OH, OH, C<u>H$_2$</u>NH), 3.04 (s, 6H, $N(CH_3)_2$), 2.29 – 2.22 (m, 2H), 2.18 (td, *J* = 7.1, 2.7 Hz, 2H, CH_2–≡), 1.95 (t, *J* = 2.7 Hz, 1H, ≡–H), 1.82 – 1.20 (m, 18H). – ^{13}C NMR (91 MHz, *CDCl$_3$*) δ 178.12, 167.74, 152.48, 128.26, 121.10, 111.08, 84.75, 71.78, 68.10, 50.20, 40.12, 38.66, 35.78, 30.51, 29.65, 29.55, 28.94, 28.69, 28.44, 27.49, 25.75, 24.86, 18.38. – HRMS (ESI) calcd. for $C_{26}H_{42}N_3O_3$ $[M+H]^+$ 444.32262, found 444.32192.

Bacterial strains and growth experiments

For all drug assays, Msm mc^{2}155 and Mtb H37Rv were grown at 37°C in Sauton's medium supplemented with 0.05% Tween 80. Per 1000 mL, Sauton's medium contains KH_2PO_4 (0.5 g), $MgSO_4$-7H_2O (0.5 g), citric acid (2.0 g), ferric ammonium citrate (0.05 g), glycerol (60 mL), asparagine (4.0 g), $ZnSO_4$ (0.1 mL of 1% solution), and NaOH (to adjust pH to 7.4).

Minimum inhibitory concentrations (MIC) of the β-lactones in Mtb and Msm were determined using the colorimetric resazurin microtiter assay as described previously. Sauton's media was dispensed in each well of a sterile flat-bottom 96 well plate. In this media, β-lactones to be tested were serially diluted two-fold across the plate. Bacteria were diluted to a calculated OD_{600} 0.003,

and equal volumes of culture were added to the media containing the serially diluted compounds. Plates were grown at 37°C for 24h (Msm) or 5d (Mtb). After the specified amount of time, 20 μL resazurin (0.05% w:v in dH_2O) was added to each well, and plates were placed back at 37°C for 12h (Msm) or 3d (Mtb). Growth and active metabolism by bacteria is reflected by the reduction of resazurin, a blue non-fluorescent dye, to resorufin, a pink highly fluorescent dye. MICs for each β-lactone were defined as the highest drug concentration that resulted in no color change upon the addition of resazurin. Prior to testing, β-lactones were tested for their ability to alter the color of resazurin, and were found to have no effect on the dye for the duration of Mtb and Msm experiments.

Minimum bactericidal concentrations of EZ120 and EZ612 in Mtb were determined by incubating various concentrations of β-lactones in 200 μL culture at an OD_{600} 0.002. After 4d, cultures were serially diluted ten-fold and plated to determine colony forming units per mL (CFU/mL).

Purification of Mtb ClpP1 and ClpP2

For all in vitro experiments, Mtb ClpP1, ClpP2, ClpP1S98A, and ClpP2S110A were purified as described previously. Briefly, subunits were individually expressed in Msm on an episomal, tetracycline (ATc)-inducible plasmid. Cultures were grown up to mid-log phase, induced with ATc (100 ng/mL) overnight, and the collected. Frozen cells (typically 5-10 g) were suspended in two volumes of buffer A (100 mM KCl, 5 mM $MgCl_2$, β-mercaptoethanol, 10% glycerol in 50 mM potassium phosphate buffer, pH 7.6) and broken by French press at 1500 p.s.i. Extracts were centrifuged at 100,000g for 1h and supernatant was mixed with 5 mL Ni-NTA agarose (Qiagen) previously equilibrated in buffer A. Resin was incubated with lysate for 4h at 4°C, and then transferred to an empty column. Proteins were eluted using a step gradient of imidazole (0, 25, 50, 100 and 200 mM) in buffer B (100 mM KCl, 5 mM $MgCl_2$, 5% glycerol in 50 mM potassium phosphate buffer, pH 7.6). Fractions containing near homogenous proteins were combined, concentrated on MWCO 10,000 cut filter (Millipore), and purified further by gel filtration on a

Sephacryl S-300 column (1.5 X 12 cm) equilibrated with buffer B. Samples were concentrated to 2-5 mg/mL and stored at -80°C.

In vitro enzyme activity assay

Activity of the ClpP1P2 complex was measured as described previously. All assays were performed at 37°C in a 96-well plate using the SpectraMax M5 Plate Reader (Molecular Devices). β-lactones were added to 0.3-3 μg ClpP1P2 (equimolar ratios of ClpP1 and ClpP2) and 5 mM activator (Z-Leu-Leu) in 80 μL of activity buffer (100 mM KCl, 5% glycerol in 50 mM potassium phosphate buffer, pH 7.6). After 10 min pre-incubation at room temperature, 50 μM fluorescent peptide substrate (Z-Gly-Gly-Leu-amc) was added to each reaction, plates were shaken for 20s, and then peptidase activities were assayed at 37°C by continuously monitoring the rate of production of fluorescent 7-amino-4-methylcoumarin (amc) from the flurogenic peptide substrate at 460 nm (Ex at 380 nm). All assays were performed in triplicate.

***In Vitro* Labeling Experiments**

Protein samples were adjusted to a final concentration of 0.04 mg protein/mL by dilution in buffer B (100 mM KCl, 5 mM $MgCl_2$, 5% glycerol in 50 mM potassium phosphate buffer, pH 7.6) prior to probe labelling. Analytical experiments were carried out in 43 μL total volume, such that once CC reagents were added, the total reaction volume was 50 μL. Reactions were initiated by addition of 150 μM probe (1 μL, 7.5 mM stock in DMSO) and allowed to incubate for 1 h at room temperature. The samples (43 μL) were then subjected to CC and SDS-PAGE analysis as described further.

***In Situ* Labeling Experiments**

For analytical and preparative *in situ* studies, bacteria were grown in Sauton's medium and harvested at an OD_{600} between 1.0 and 3.0 by centrifugation at 6000 g. 3 mL of culture were harvested for analytical and 15 mL for preparative studies, respectively. The cells were washed with PBS (1 mL) and resuspended in 100 and 500 μL of PBS for analytical and preparative

experiments. Unless indicated otherwise, bacteria were incubated for 2 h with 150 µM probe at room temperature for analytical preparative studies. Subsequently, the cells were washed with PBS (3 x 1 mL) and lysed by bead beating (MP Biomedicals, 4 x 30 sec, 5min on ice between each round) in 100 and 500 µL of PBS for analytical and preparative experiments respectively. The proteomes were subjected to CC and SDS-PAGE analysis as described further.

Click Chemistry (CC) and Target Identification

Reactions for preparative enrichment were carried out together with a control lacking the probe to subtract unspecific background protein binding on avidin-agarose beads in the final MS results of the biotin-avidin-enriched samples. To 43 µL of the probe pre-incubated protein/proteome samples were added the reporter-tagged azide reagents rhodamine-azide (100 µM, 1 µL of 5 mM stock in DMSO) for analytical or rhodamine-biotin-azide (50 µM, 2.5 µL of 10 mM stock in DMSO) for preparative scale, followed by tris(2-carboxyethyl)phosphine hydrochloride (1 mM, 1 µL for analytical or 0.5 mM, 5 µL for preparative scale of 52 mM stock in H_2O) and as ligand tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl] amine (100 µM, 3 µL for analytical or 50 µM, 15 µL for preparative scale of 1.7 mM stock in *tert*-BuOH/DMSO 4.4:1). Samples were gently vortexed, and the 1,3-dipolar cycloaddition was initiated by the addition of $CuSO_4$ (1 mM, 1 µL for analytical or 0.5 mM, 5 µL for preparative scale of 50 mM stock in H_2O). The reactions were incubated at room temperature for 1 h[13,20,23]. For analytical gel electrophoresis (SDS-PAGE), 50 µL of 2×SDS loading buffer was added, the samples were heated at 96 °C for 20 min for Mtb samples and 6 min for Msm samples, and 50 µL were applied on the gel. For preparative scale proteins were precipitated after CC using an equal volume of pre-chilled acetone (500 µL). The samples were then stored on ice for 20 min and centrifuged at 13000 rpm for 20 min. The supernatant was discarded, and the pellet was washed with pre-chilled methanol (2 x 200 µL) and resuspended by sonication (Bandelin Sonopuls, 5 sec, 20%). Subsequently, the pellet was dissolved in 0.4% SDS in PBS (1 mL) by sonication and incubated under gentle mixing with 50 µL of washed avidin-agarose beads (Sigma-Aldrich) for 1 h at room temperature. The beads were washed with 0.4%

SDS in PBS (3 x 1 mL), with urea (2 x 1 mL, 6 M) and with PBS (3 x 1 mL). 50 µL of 2×SDS loading buffer were added to the beads and the proteins released for preparative SDS-PAGE by 20 min incubation for Mtb samples and 6 min incubation for Msm samples at 96 °C. Fluorescence was recorded in a Fujifilm Las-4000 luminescent image analyzer with a Fujinon VRF43LMD3 lens and a 575DF20 filter. Gel bands were isolated, washed, and tryptically digested as described previously[24].

Mass Spectrometry and Bioinformatics

Tryptic peptides were loaded onto a Dionex C18 Nano Trap column (100 µm) and subsequently eluted and separated by a Dionex C18 PepMap 100 (3 µm) column for analysis by tandem MS, followed by high resolution MS using a coupled Dionex Ultimate 3000 LC Thermo Finnegan LTQ-FT MS system. The mass spectrometry data were searched using the SEQUEST algorithm against the corresponding databases via the software "Proteome Discoverer 1.3". The search was limited to only tryptic peptides, two missed cleavage sites, monoisotopic precursor ions, and a peptide tolerance of <10 ppm. Filters were set to further refine the search results. The *X*corr vs charge state filter was set to *X*corr values of 1.5, 2.0, and 2.5 for charge states 1+, 2+, and 3+, respectively. The number of different peptides has to be ≥2, and the peptide probability filter was set to <0.001. These filter values are similar to others previously reported for SEQUEST analysis[25]. Maximum X_{corr} values of each run as well as the total number of obtained peptides are reported in **FIGURE 4.5D**. Comparison of labelling with genetic properties and characteristics of the species was conducted by BLAST search for homologous sequences in the strains.

Binding Site Identification

To discover the binding site of ClpP1 and ClpP2, a 1:1 mixture of both enzymes as well as a 1:1 mixture of the active serine mutants (ClpP1-S98A, ClpP2-S110A) was incubated with probe P1 and after chymotryptic digestion the peptide fragments were analyzed by MS. First, 0.68 µg of Protein in 17 µL buffer B (100 mM KCl, 5 mM $MgCl_2$, 5% glycerol in 50 mM potassium phosphate

buffer, pH 7.6) was incubated for 1 h at room temperature with 150 µM of **P1**. The buffer was exchanged and adjusted to 25 mM NH_4HCO_3, giving a total volume of 100 µL. For digestion with chymotrypsin (sequencing grade, Roche Diagnostics) 1 mM $CaCl_2$ was added prior and after addition of 250 ng chymotrypsin the samples were incubated for 18 h at room temperature. The samples were dried, solved in 100 µL 0.1% formic acid and analyzed by LC-MS with MS-MS and high-resolution mass spectra. The spectra of the modified peptides are shown in **FIGURE 4.3**.

Determination of cytotoxicity / MTT assay

The cytotoxicity of the compound **EZ120** against mouse macrophages (J774) was measured using the MTT test (Sigma-Aldrich). The principle of the MTT assay consists in the reduction of the tetrazolium compound MTT by metabolically active cells to insoluble purple formazan dye crystals as a measure of cells viability. Cells were cultured in DMEM low glucose supplemented with 10% FCS gold (PAA). Cells from subconfluent cultures were used for the assay. In a 96well flat-bottom plate (Nunclon) $1 \cdot 10^4$ cells were plated in 100 µL medium and cultured for 24 h to obtain 40% confluent cultures. The compound was diluted 1:100 from DMSO stocks of various concentrations in 100 µL of culture medium and added to the cells after careful removal of the blank medium. After 24 h incubation, 20 µL of MTT substrate solution (5 mg/ml in PBS) were added and the cells were incubated for 2 h. The medium was discarded and the cells were lysed in 200 µL DMSO. The complete dissolution of the formazan salt was determined and the optical density measured at 570 nm (background subtraction at 630 nm). The assay was performed in triplicates.

WORKS CITED

1. World Health Organization Global tuberculosis control: WHO report 2011. Geneva, Switzerland: World Health Organization; 2011. (2011).

2. Multidrug, W. extensively drug-resistant TB (M/XDRTB): 2010 global report on surveillance and response. Publication no. (2010).

3. Velayati, A. A. *et al.* Emergence of New Forms of Totally Drug-Resistant Tuberculosis

Bacilli: Super Extensively Drug-Resistant Tuberculosis or Totally Drug-Resistant Strains in Iran. *Chest* **136**, 420–425 (2009).

4. Udwadia, Z. F., Amale, R. A., Ajbani, K. K. & Rodrigues, C. Totally drug-resistant tuberculosis in India. *CLIN INFECT DIS* **54**, 579–581 (2012).

5. MD, D. R. R. K., MD, S. V., MD, N. S., MD, N. D. & MD, P. H. M. B. Surgical treatment of drug-resistant tuberculosis. *Lancet Infect Dis* **12**, 157–166 (2012).

6. Ginsberg, A. M. Drugs in Development for Tuberculosis. *Drugs* **70**, 2201–2214 (2010).

7. Böttcher, T. & Sieber, S. A. Beta-lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of Staphylococcus aureus. *J Am Chem Soc* **130**, 14400–14401 (2008).

8. Frees, D., Sørensen, K. & Ingmer, H. Global virulence regulation in Staphylococcus aureus: pinpointing the roles of ClpP and ClpX in the sar/agr regulatory network. *Infection and Immunity* **73**, 8100–8108 (2005).

9. Böttcher, T. & Sieber, S. A. Structurally refined beta-lactones as potent inhibitors of devastating bacterial virulence factors. *ChemBioChem* **10**, 663–666 (2009).

10. Kress, W., Maglica, Z. & Weber-Ban, E. Clp chaperone-proteases: structure and function. *Res Microbiol* **160**, 618–628 (2009).

11. Raju, R. M. *et al.* Mycobacterium tuberculosis ClpP1 and ClpP2 Function Together in Protein Degradation and Are Required for Viability in vitro and During Infection. *PLoS Pathog* **8**, e1002511 (2012).

12. Akopian, T. *et al.* The active ClpP protease from M. tuberculosis is a complex composed of a heptameric ClpP1 and a ClpP2 ring. *EMBO J* 1–13 (2012).doi:10.1038/emboj.2012.5

13. Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-Based Protein Profiling in Vivo Using a Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *J Am Chem Soc* **125**, 4686–4687 (2003).

14. Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29–40 (2007).

15. Brennan, P. J. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. *Tuberculosis* **83**, 91–97 (2003).

16. Groll, M., Balskus, E. P. & Jacobsen, E. N. Structural Analysis of Spiro β-Lactone Proteasome Inhibitors. *J Am Chem Soc* **130**, 14981–14983 (2008).

17. Gandotra, S., Lebron, M. B. & Ehrt, S. The Mycobacterium tuberculosis proteasome active site threonine is essential for persistence yet dispensable for replication and resistance to nitric oxide. *PLoS Pathog* **6**, (2010).

18. Zeiler, E., Korotkov, V. S., Lorenz-Baath, K., Böttcher, T. & Sieber, S. A. Development and characterization of improved β-lactone-based anti-virulence drugs targeting ClpP. *Bioorganic & Medicinal Chemistry* **20**, 583–591 (2012).

19. Rathore, S. *et al.* A cyanobacterial serine protease of Plasmodium falciparum is targeted to the apicoplast and plays an important role in its growth and development. *Molecular Microbiology* no–no (2010).doi:10.1111/j.1365-2958.2010.07251.x

20. Böttcher, T. & Sieber, S. A. Beta-lactones as privileged structures for the active-site labeling of versatile bacterial enzyme classes. *Angewandte Chemie (International ed in English)* **47**, 4600–4603 (2008).

21. Danheiser, R. L. & Nowick, J. S. A practical and efficient method for the synthesis of. beta.-lactones. *The Journal of Organic chemistry* **56**, 1176–1185 (1991).

22. Zemski Berry, K. A., Turner, W. W., VanNieuwenhze, M. S. & Murphy, R. C. Stable isotope labeled 4-(dimethylamino)benzoic acid derivatives of glycerophosphoethanolamine lipids. *Anal Chem* **81**, 6633–6640 (2009).

23. Böttcher, T. & Sieber, S. A. Showdomycin as a Versatile Chemical Tool for the Detection of Pathogenesis-Associated Enzymes in Bacteria. *J Am Chem Soc* **132**, 6964–6972 (2010).

24. Sieber, S. A., Niessen, S., Hoover, H. S. & Cravatt, B. F. Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat Chem Biol* **2**, 274–281 (2006).

25. Mirza, S. P., Halligan, B. D., Greene, A. S. & Olivier, M. Improved method for the analysis of membrane proteins by mass spectrometry. *Physiological Genomics* **30**, 89–94 (2007).

Chapter 5:

Conclusions and Future Perspectives

Section 5.1 – Clp protease as a critical regulator of growth and adaptation in mycobacteria

It was initial inquiries into determining how bacteria dealt with abnormal proteins that led to the discovery and characterization of ATP-dependent protein turnover[1]. Since that time, the array of proteolytic complexes responsible for this turnover in bacteria has been extremely well characterized from a structural and biochemical perspective[2,3]. Numerous researchers have focused on developing an intricate picture of how these enzymes maintain quality control in the cell through the degradation of mislocalized, misfolded, and abnormal protein products[4]. What has become increasingly clear, however, is that the proteases also play a critical role in the coordination of cellular activity through the proteolysis of endogenous proteins. Many studies have shown an association (either directly or through mutational pathway analysis) between the protease and a particular protein, but deciphering the functional consequence of targeted protein degradation has been far more challenging[5,6]. It is often difficult to determine whether degradation of a particular protein is a passive process or an active one that is required for regulation. As numerous proteolytic complexes appear to be required for normal growth of *Mycobacterium tuberculosis* (Mtb), a divergence from most bacteria, the field of protease biology may have found an unlikely model organism in the slow-growing, persistent pathogen that causes nearly 1.3 million deaths annually[7]. Depletion of these enzymes and accumulation of their substrates in Mtb should have a visible growth effect, and so identifying the proteins they degrade could uncover pathways that must be critically regulated through targeted protein turnover.

It was in this ambitious hope that my thesis work was forged. We chose to work on the Clp proteolytic complex in mycobacteria for several reasons. First, genome wide-screens in Mtb suggested that all components of the complex (proteolytic core and ATPase adapter proteins) were absolutely required for growth[8]. This was strikingly different from most model organisms, like *Escherichia coli* and *Bacillus subtilis*, where Clp has been extensively studied but found to be

dispensable for vegetative growth[9,10]. Second, the genetic structure of Clp appeared to be different in mycobacteria, which appeared to encode two proteolytic subunit genes (clpP1 and clpP2) in a single operon, instead of the canonical single proteolytic subunit (clpP). This led to the question of whether Mtb had two functional Clp proteases or one enzyme made up of multiple proteolytic subunits.

Partnering with skilled collaborators in the laboratory of Alfred Goldberg, where Clp was first discovered in *E. coli* in 1980s[11], we epitope-tagged ClpP1 and ClpP2, showing that the proteins interacted intimately in vivo. Purifying the subunits and reconstituting activity in an in vitro peptidase assay further demonstrated that equimolar amounts of ClpP1 and ClpP2 were required for efficient proteolysis, and that neither subunit could support proteolysis individually. Together, this data suggested that ClpP1 and ClpP2 formed a single, proteolytic heterotetradecameric core in mycobacteria, denoted ClpP1P2. To validate the importance of these genes in vivo, I constructed conditional Clp protease mutants in *Mycobacterium smegmatis* (Msm). Both clpP1 and clpP2 were independently required for growth, and deletion of either subunit resulted in rapid bacterial death. As depletion of a gene rarely leads to a bactericidal effect, this finding further stressed the central role of the mycobacterial Clp protease. In order to disrupt the ClpP1P2 proteolytic core in Mtb, we constructed active site mutants of ClpP1, in which the serine responsible for peptide bond hydrolysis was mutated to an alanine. Expression of these mutant subunits had a visible (but slight) dominant negative effect on growth in vitro, presumably by displacing wildtype ClpP1 subunits from the heteromeric complex and reducing the efficiency of proteolysis. In contrast to the minor dominant negative effect in vitro, expression of the active site mutants profoundly reduced in vivo survival of Mtb in a mouse model of infection, leading us to believe that Clp protease has additional roles during infection and could be an important effector of virulence.

Having validated the stringent requirements of Clp protease, we next turned towards understanding the functional significance of the complex in mycobacteria. In *E. coli*, Clp is involved in the ribosome rescue pathway, through the degradation of SsrA-tagged peptides that are released from stalled ribosomes by the translation of a tmRNA substrate[12]. Using a conditional Msm mutant, in which the ClpP2 protein could be inducibly degraded, I showed that this general feature is conserved in mycobacteria. Clp appears to be generally important in maintaining protein quality control in the cell, as cells with suboptimal amounts of the protease were highly susceptible to antibiotics that induce mistranslation but showed no less susceptibility, when compared to wildtype bacteria, to agents that simply halt translation.

Because the quality control functions of Clp protease are conserved across most bacteria, these studies did not necessarily move us towards understanding why Clp was divergently essential in mycobacteria. We hypothesized that the requirements of Clp protease for normal growth were tied to the endogenous proteins it was degrading. For example, if Clp degraded a toxic protein or a repressor of normal growth, then the accumulation of those substrates that occurred upon depletion of the protease would have a deleterious effect on the cell. This hypothesis was difficult to test, because we were not probing the function of an important protein when it was present, but rather the necessity of certain proteins to be absent, or actively removed from the cell. In order to identify the set of proteins whose Clp-dependent degradation might be important for cell survival, we turned to proteomic profiling of conditional Clp mutants in both Mtb and Msm, which revealed a slew of proteins that were over-represented in bacterial populations that lacked wildtype levels of the enzyme. Through a set of validation experiments, we provided multiple lines of evidence that the transcriptional repressor WhiB1 was a substrate of Clp protease.

This observation was intriguing for multiple reasons. First, WhiB1 is an essential DNA-binding protein that is known to repress it's own transcription, and thought to act as a repressor at multiple operators[13]. Thus, we thought the essentiality of Clp protease could potentially be

explained by its action on WhiB1, if elevated amounts of the repressor suppress transcription of these essential loci to a level that inhibits growth. Accordingly, blocking of Clp recognition of WhiB1 resulted in a degradation-deficient WhiB1 allele that was ultra-stable and toxic. Though hard to eliminate non-specific causes of fusion protein toxicity, we demonstrated that this allele was still functional as a transcriptional repressor and expression still exhibited a fitness cost even at physiological levels of protein expression. Second, deregulation of WhiB1 proteolysis could not be overcome by transcriptional control, suggesting that protein turnover was at least a co-dominant modality of WhiB1 regulation. It is interesting that both depletion and stabilization of WhiB1 result in bacterial death, and in this light, it would make sense that there would be multiple regulatory mechanisms to keep levels of WhiB1 tightly regulated. There are only a few examples where proteolysis has been shown to be the primary regulatory mechanism of an essential process, but here we have shown that turnover of WhiB1 by Clp protease is required, and in the absence of effective proteolysis, cell death ensues.

While we have provided evidence for the importance of Clp-dependent WhiB1 degradation, a major unanswered question is why is WhiB1 turnover critically required for growth? While we hypothesized that stabilization of WhiB1 may overly repress the transcription of essential genes, this must be confirmed by transcriptional analysis of the degradation-deficient allele. Extremely preliminary studies that I have conducted suggest that the picture may be more complicated than presented. Taking WhiB1 ChIP-Seq data deposited in the Tuberculosis Database (www.tbdb.org), which reported 72 genes regulated by WhiB1 in Mtb, I found Msm homologs for the thirteen essential Mtb genes in the putative regulon[14]. Quantitative PCR analysis in Msm strains over-expressing GFP-WhiB1 (the rapidly degraded allele) and WhiB1-GFP (the stabilized allele) revealed a very jumbled picture **(APPENDIX 4)**. In some cases, repression of the essential loci was equal for both fusions, while with other genes there was actually transcriptional activation by WhiB1-GFP. This initial inquiry is flawed for many reasons (ie. using preliminary ChIP-Seq data, validating ChIP-Seq in a different species), but it does make one wonder that perhaps the toxicity

of WhiB1-GFP may be more complex than excessive repression of essential genes in the canonical WhiB1 regulon. To this end, it would be interesting to conduct ChIP-Seq analysis on the stabilized allele to see if elevated concentrations of WhiB1 result in additional DNA-binding sites. Ideally, the lethality of WhiB1 stabilization will be explained by highlighting the transcriptional effects of having excessive concentrations of the essential transcription factor in the cell.

WhiB1 is just one potential substrate revealed from screening, but the combination of qPCR and fusion protein analysis has highlighted an efficient method of validating proteins degraded by Clp protease that were identified from our initial proteomic screens. Using this methodology, we have also shown that CarD and RpL28 are likely Clp substrates. Though blocking degradation of these proteins failed to have an observable effect on growth, turnover of these proteins may still be an important regulatory mechanism in mycobacteria. We hope to apply this validation framework to a larger set of the 132 Mtb and 107 Msm proteins identified as putative Clp substrates through proteomic screening. While gene ontology analysis failed to reveal the enrichment of any particular pathways among the over-represented proteins in Clp-deficient Mtb, there were numerous essential hypothetical proteins and uncharacterized DNA-binding proteins identified. Determining if these effectors are true substrates, and subsequently analyzing the cellular effects of stabilized allelic forms, will highlight proteins and pathways that are critically regulated through targeted degradation.

In reality, however, there is unlikely one “silver bullet” substrate that explains why Clp protease is essential in mycobacteria. Evidence for this is derived from whole-genome screens for essential genes, which found both the ATPase adapters, ClpX and ClpC1, to also be required for normal growth[8]. As these adapters dock with the ClpP1P2 proteolytic core and govern substrate recognition, their essentiality suggests that they independently regulate the degradation of different sets of substrates. It may be that the integration of inefficient proteolysis of numerous proteins explains the growth inhibition that results from Clp depletion.

Our work has simply scratched the surface in understanding the importance of the Clp protease complex in mycobacteria. As is the nature of science, countless questions remain. Given the essentiality of the ATPase adapters, how substrates are recognized for degradation is another interesting question that remains to be solved. Our work to date has focused on the essentiality and biology surrounding the proteolytic core of Clp, ClpP1P2. However, this core is simply a blind chamber that is governed by ATPase adapters, ClpX and ClpC1, which are responsible for substrate recognition and providing the energy for translocation of proteins into the proteolytic chamber[15]. Our lab is currently constructing conditional mutants of both ATPase adapters in Mtb. Once developed, these strains should be subjected to similar proteomic analysis to determine the set of proteins whose turnover is regulated specifically by each of the adapters.

Observations from model organisms suggest that the adapters may coordinate distinct biology. In *B. subtilis*, the ClpCP complex shuts down central biosynthetic pathways by degrading rate-limiting enzymes responsible for cell wall (GlmS, MurAA), branched chain amino acid (IlvB), purine (PurF), and pyrimidine (PyrB) synthesis[16]. While the functional significance of this turnover has yet to be shown, one could imagine that ClpCP might be critically important in the transition of *B. subtilis* from nutrient-rich to nutrient-depleted environments. In contrast, ClpXP in *E. coli* has been shown to degrade the enzymes isocitrate lyase and malate synthase, and the DNA crystallization protein (Dps), which are required for survival during starvation, and would need to be degraded upon shifting of *E. coli* from nutrient-depleted to nutrient rich environments[5,17]. In mycobacteria, the ClpP1P2 proteolytic core may, likewise, be the common endpoint of numerous proteins whose fates are actually determined through regulation at the ATPase adapter level. In this model, the ATPases are the keys to understanding how Clp regulates the degradation of particular substrates and controls various cellular activities.

Understanding how the ATPases govern protein recognition and facilitate proteolysis would be greatly enhanced with the reconstitution of substrate degradation in vitro. Currently, we have developed ClpC1P1P2 peptidase assays, but have yet to test the degradation of any putative Clp substrates. The purification of ClpX has proven challenging, and considerable effort must be invested into obtaining a functional ClpXP1P2 enzyme complex in vitro. Demonstration of in vitro degradation by one of the two complexes would be the most direct way to validate substrates of Clp protease. In addition, once these assays are up and running, the testing of substrate degradation could also help to reveal additional co-factors and adapters that might be required for recognition by Clp. In *E. coli*, for example, the small SspB protein is required to tether SsrA-tagged peptides to ClpXP for degradation[18]. Despite lacking an SspB homolog, we have shown that mycobacteria are still able to degrade SsrA-tagged substrates. In vitro testing of GFP-SsrA degradation by the ClpXP1P2 complex could ascertain whether mycobacteria simply did away with SspB or employ a non-homologous small protein adapter to facilitate SsrA recognition.

While the Clp complexes are clearly important regulators of cellular activity, they too must also be regulated. Several groups have shown that the positive regulator ClgR induces expression of clpP1, clpP2, and clpX[19]. Interestingly, clgR is highly induced upon reaeration of a hypoxic Mtb culture, suggesting that the ClpXP1P2 protease may be critically required in rapidly facilitating adaptation to oxygen-replete environments[20]. Coupling such transcriptional analysis, with protein turnover analysis to see which proteins are destabilized upon reaeration could highlight the critical substrates that must be degraded to reprogram the cells. Beyond this, though, we know next to nothing about how Clp is regulated. Studying the induction or repression of clpP1, clpP2, clpC1, and clpX transcript in various conditions or growth phases would be a good place to start, offering us a view of critical time points or environments in which Clp activity is called upon by mycobacteria.

Section 5.2 – Clp protease as a chemotherapy target for the treatment of tuberculosis

As we continue to unearth the important biological processes regulated by the essential Clp protease, it is critical that we harness this knowledge to aid in the fight against *Mycobacterium tuberculosis*. The rapid bacterial death that ensued from genetic depletion of Clp protease excited us tremendously, as it validated the enzyme as an ideal drug target, and suggested that chemical inhibition could be effective for treatment. Fortuitously, we came across a class of compounds, β-lactones, that had been described as Clp-specific inhibitors in both *Staphylococcus aureus* and *Listeria monocytogenes*[21,22]. In these pathogens, Clp is dispensable for growth in vitro, but is absolutely required for virulence during infection. For this reason, the compounds were being pursued as anti-virulence agents in *S. aureus* and *L. monocytogenes*, but we wanted to evaluate their potential as traditional growth inhibitors in *Mycobacterium tuberculosis*. One β-lactone, EZ120, had a superb killing profile, with an MIC of 1.6 μM and an MBC between 3-6 μM. Furthermore, we were able to show that EZ120 inhibited ClpP1P2 proteolysis in an in vitro purified protein assay, and that the compound interacted with the enzyme complex in intact cells. Together, this evidence validates the β-lactones as a potential class of antimycobacterial compounds and Clp protease as a suitable target for more focused drug development.

Extremely preliminary studies in eukaryotic macrophages suggests that EZ120 has a therapeutic index of 200:1, but many issues still surround the development of this compound and the broader class of β-lactones as suitable, lead candidates for drug development. For example, they have an extremely short half-life in plasma, and are poorly soluble in aqueous media. Our collaborators in the laboratory of Stephan Sieber are working to create libraries of more pharmaceutically attractive compounds, and it will be important to test these new chemical entities for bactericidal activity and Clp inhibition.

On other fronts, we have also preliminary tested a set of peptidomimetic compounds from the laboratory of Alfred Goldberg, and shown that they that inhibit ClpP1P2 proteolysis in vitro and the growth of Mtb and Msm. As velcade, a peptidomimemtic eukaryotic proteasome inhibitor used in the treatment of multiple myeloma, is already approved and in clinical use, the path to development of the boronate compounds may be less tortuous than the β-lactones[23]. However, these compounds come with their own slew of issues, namely ensuring that they do not have significant toxicity through inhibition of eukaryotic proteases.

Given that descriptions of total-drug resistant tuberculosis are now cropping up in the literature, and the drug pipeline for tuberculosis is disparagingly barren, we must ramp up efforts to exploit essential pathways in mycobacteria. While we have demonstrated the inhibitory activity of a few compounds, presumably via Clp protease inhibition, significant work is still needed to take these compounds from the initial discovery phase to the generation of ideal lead compounds for antibiotic drug development.

Section 5.3 – Why are the proteolytic complexes so critical for survival in mycobacteria?

While this thesis has focused on the biological and therapeutic significance of the Clp protease, it is interesting to consider, in closing, why mycobacteria would have evolved to rely so heavily on protein turnover. As stated above, genome wide screens for essential genes in mycobacteria have found that multiple proteolytic complexes (Clp, FtsH, HtrA) are required for growth, a surprising discovery given that these enzymes are mostly dispensable for vegetative growth in most other organisms where they have been studied. Though territory full of conjecture, we can posit several hypotheses as to why the class in general may be divergently important in mycobacteria.

The first is the “stress” hypothesis. While it is true that the proteolytic complexes are dispensable for normal growth in most organisms, they become critically required in almost all cases during more stringent growth conditions. Using Clp as an example, the protease is absolutely required for survival during infection in nearly every pathogenic organism where it has been studied[24-26]. And for the non-pathogenic organisms, Clp is required for survival in most forms of in vitro stress, and depletion of the protease is associated with impaired transition between certain growth phases[27]. Taking this into account, it could just be that mycobacteria are more sensitive to stress, and that “normal” growth actually requires continual adaptation carried out by the proteolytic complexes. Alternatively, the ability of Mtb to survive for decades in harsh, competitive niches may have necessitated an expansion of targeted proteolysis as a regulatory mechanism. In support of this, one preliminary experiment I ran during my doctoral work suggested that the E. coli Clp machinery did not degrade WhiB1. This suggests that different rules may govern the recognition of proteins by Clp in mycobacteria. Especially given that they may apply to WhiB1 degradation, these differential recognition patterns may be linked to the expanded role of Clp and explain why the protease is divergently essential in mycobacteria. Survival of Mtb in stress-filled environments, like the human macrophage, may also require adaptation to be rapid in order to prevent death. The fact that ClgR expression and activation of various proteases are required for the intracellular parasitism of Mtb[28] suggests that proteolysis is a key regulatory feature in the adaptation of Mtb. The essentiality of these enzymes may have arisen from the evolutionary benefit of rapid adaptation afforded by protein turnover, as opposed to slower transcriptional modulation.

The second possible explanation of why multiple proteases are divergently essential in mycobacteria is the “slow grower” hypothesis. With a replication rate of 20 minutes, *E. coli* is quickly able to dilute the presence of toxic proteins or growth repressors through repeated rounds of division. Mtb, on the other hand, divides once every 1,440 minutes, nearly 2 orders of magnitude slower. The presence of harmful proteins in Mtb may have a much more deleterious

effect on the bacteria simply because their effective concentrations are much higher for far longer. Thus, the slow growth rate of mycobacteria may have necessitated the development of active enzymatic mechanisms to deal with proteins (both regulatory and toxic) that inhibit growth, while the fast growing bacteria may have just relied on divisional dilution for the same effect.

Regardless of the reason why, the essentiality of these enzymes gives us an interesting entrée into understanding how protein turnover is employed as a regulatory feature in mycobacteria, and prokaryotes in general. As a young biology student, I remember learning the central dogma and never questioning what happened to proteins when they were no longer needed. In hindsight, as I quickly learned through my doctoral work, a transcriptional-centric view of the central dogma highlights important regulatory features, but fails to address where proteins go to die when they have outlived their utility, or what cells must do when a response must be initiated within seconds of encountering an environmental stress. We have just begun to probe these mechanisms in mycobacteria, but their elucidation will undoubtedly lead us to novel biology, and facilitate a more holistic understanding of how bacteria regulate essential cellular processes.

WORKS CITED

1. Murakami, K., Voellmy, R. & Goldberg, A. L. Protein degradation is stimulated by ATP in extracts of Escherichia coli. *J Biol Chem* **254**, 8194–8200 (1979).

2. Schmidt, M., Lupas, A. N. & Finley, D. Structure and mechanism of ATP-dependent proteases. *Curr Opin Chem Biol* **3**, 584–591 (1999).

3. Sauer, R. T. & Baker, T. A. AAA+ Proteases: ATP-Fueled Machines of Protein Destruction. *Annu Rev Biochem* (2010).doi:10.1146/annurev-biochem-060408-172623

4. Goldberg, A. L. Degradation of abnormal proteins in Escherichia coli (protein breakdown-protein structure-mistranslation-amino acid analogs-puromycin). *Proc Natl Acad Sci USA* **69**, 422–426 (1972).

5. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. & Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol Cell* **11**, 671–683 (2003).

6. Hong, S.-J., Lessner, F. H., Mahen, E. M. & Keiler, K. C. Proteomic identification of tmRNA substrates. *Proc Natl Acad Sci USA* **104**, 17128–17133 (2007).

7. World Health Organization Global tuberculosis control: WHO report 2011. Geneva,

Switzerland: World Health Organization; 2011. (2011).

8. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology* **48**, 77–84 (2003).

9. Damerau, K. & St John, A. C. Role of Clp protease subunits in degradation of carbon starvation proteins in Escherichia coli. *J Bacteriol* **175**, 53–63 (1993).

10. Gerth, U., Kruger, E., Derré, I., Msadek, T. & Hecker, M. Stress induction of the Bacillus subtilis clpP gene encoding a homologue of the proteolytic component of the Clp protease and the involvement of ClpP and ClpX in stress tolerance. *Molecular Microbiology* **28**, 787–802 (1998).

11. Hwang, B. J., Woo, K. M., Goldberg, A. L. & Chung, C. H. Protease Ti, a new ATP-dependent protease in Escherichia coli, contains protein-activated ATPase and proteolytic functions in distinct subunits. *J Biol Chem* **263**, 8727–8734 (1988).

12. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* **12**, 1338–1347 (1998).

13. Smith, L. J. *et al.* Mycobacterium tuberculosis WhiB1 is an essential DNA-binding protein with a nitric oxide-sensitive iron-sulfur cluster. *Biochem J* **432**, 417–427 (2010).

14. Reddy, T. B. K. *et al.* TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res* **37**, D499–D508 (2009).

15. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**, 27–43 (1999).

16. Gerth, U. *et al.* Clp-dependent proteolysis down-regulates central metabolic pathways in glucose-starved Bacillus subtilis. *J Bacteriol* **190**, 321–331 (2008).

17. Stephani, K., Weichart, D. & Hengge, R. Dynamic control of Dps protein levels by ClpXP and ClpAP proteases in Escherichia coli. *Molecular Microbiology* **49**, 1605–1614 (2003).

18. Hersch, G. L., Baker, T. A. & Sauer, R. T. SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags. *Proc Natl Acad Sci USA* **101**, 12136–12141 (2004).

19. Bellier, A. & Mazodier, P. ClgR, a novel regulator of clp and lon expression in Streptomyces. *J Bacteriol* **186**, 3238–3248 (2004).

20. Sherrid, A. M., Rustad, T. R., Cangelosi, G. A. & Sherman, D. R. Characterization of a Clp Protease Gene Regulator and the Reaeration Response in Mycobacterium tuberculosis. *PLoS ONE* **5**, e11622 (2010).

21. Böttcher, T. & Sieber, S. A. Beta-lactones as specific inhibitors of ClpP attenuate the production of extracellular virulence factors of Staphylococcus aureus. *J Am Chem Soc* **130**, 14400–14401 (2008).

22. Böttcher, T. & Sieber, S. A. Beta-lactones decrease the intracellular virulence of Listeria

monocytogenes in macrophages. *ChemMedChem* **4**, 1260–1263 (2009).

23. Goldberg, A. Bortezomib's Scientific Origins and Its Tortuous Path to the Clinic. *Bortezomib in the Treatment of Multiple Myeloma* (2011).

24. Gaillot, O., Pellegrini, E., Bregenholt, S., Nair, S. & Berche, P. The ClpP serine protease is essential for the intracellular parasitism and virulence of Listeria monocytogenes. *Molecular Microbiology* **35**, 1286–1294 (2000).

25. Frees, D., Sørensen, K. & Ingmer, H. Global virulence regulation in Staphylococcus aureus: pinpointing the roles of ClpP and ClpX in the sar/agr regulatory network. *Infection and Immunity* **73**, 8100–8108 (2005).

26. Ingmer, H. & Brøndsted, L. Proteases in bacterial pathogenesis. *Res Microbiol* **160**, 704–710 (2009).

27. Frees, D., Savijoki, K., Varmanen, P. & Ingmer, H. Clp ATPases and ClpP proteolytic complexes regulate vital biological processes in low GC, Gram-positive bacteria. *Molecular Microbiology* **63**, 1285–1295 (2007).

28. Estorninho, M. *et al.* ClgR regulation of chaperone and protease systems is essential for Mycobacterium tuberculosis parasitism of the macrophage. *Microbiology* (2010).doi:10.1099/mic.0.042275-0

Appendices

Appendix 1 – The active ClpP protease from *M. tuberculosis* is a complex composed of a heptameric ClpP1 and a ClpP2 ring

The EMBO Journal (2012), 1–13 | © 2012 European Molecular Biology Organization | All Rights Reserved 0261-4189/12
www.embojournal.org

THE EMBO JOURNAL

The active ClpP protease from *M. tuberculosis* is a complex composed of a heptameric ClpP1 and a ClpP2 ring

Tatos Akopian[1,3], Olga Kandror[1,3], Ravikiran M Raju[2], Meera UnniKrishnan[2], Eric J Rubin[2] and Alfred L Goldberg[1,*]

[1]Department of Cell Biology, Harvard Medical School, Boston, MA, USA and [2]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

Mycobacterium tuberculosis **(Mtb) contains two *clpP* genes, both of which are essential for viability. We expressed and purified Mtb ClpP1 and ClpP2 separately. Although each formed a tetradecameric structure and was processed, they lacked proteolytic activity. We could, however, reconstitute an active, mixed ClpP1P2 complex after identifying N-blocked dipeptides that stimulate dramatically (>1000-fold) ClpP1P2 activity against certain peptides and proteins. These activators function cooperatively to induce the dissociation of ClpP1 and ClpP2 tetradecamers into heptameric rings, which then re-associate to form the active ClpP1P2 2-ring mixed complex. No analogous small molecule-induced enzyme activation mechanism involving dissociation and re-association of multimeric rings has been described. ClpP1P2 possesses chymotrypsin and caspase-like activities, and ClpP1 and ClpP2 differ in cleavage preferences. The regulatory ATPase ClpC1 was purified and shown to increase hydrolysis of proteins by ClpP1P2, but not peptides. ClpC1 did not activate ClpP1 or ClpP2 homotetradecamers and stimulated ClpP1P2 only when both ATP and a dipeptide activator were present. ClpP1P2 activity, its unusual activation mechanism and ClpC1 ATPase represent attractive drug targets to combat tuberculosis.**

The EMBO Journal advance online publication, 27 January 2012; doi:10.1038/emboj.2012.5
Subject Categories: proteins; microbiology & pathogens
Keywords: cleavage preferences; dipeptide activator; protease ClpP1P2

Introduction

Tuberculosis is a devastating disease that affects worldwide about 100 million people and causes nearly 2 million deaths annually. It has been estimated that one third of all humans is infected with latent Mycobacterium tuberculosis (Mtb). Moreover, Mtb has become increasingly resistant to available antibiotics. Consequently, it is important to identify and characterize new therapeutic targets in Mtb and to synthesize selective inhibitors. Ideal targets for drug development should be enzymes essential for bacterial viability that differ in physicochemical properties and specificity from those present in humans. ClpP1 has been recently validated to be essential in Mtb (Ollinger *et al*, 2012) and in related studies, we established that the proteases encoded by *clpP1* and by *clpP2* are required both for the growth of Mtb and for its virulence during murine infection (Sassetti *et al*, 2003; Raju *et al*, 2011). Therefore, ClpP1 and ClpP2 are highly attractive drug targets, especially since they are not present in the cytosol of mammalian cells (where protein breakdown occurs primarily by the ubiquitin proteasome pathway or by lysosomes), and these enzymes differ markedly from the mitochondrial Clp complex.

ClpP is a highly conserved, multimeric serine protease originally discovered (Hwang *et al*, 1987; Katayama-Fujimura *et al*, 1987) and extensively characterized in *E. coli* (Maurizi *et al*, 1990b, 1994, 1998; Yu and Houry, 2007). ClpP homologues exist in a wide range of bacteria, as well as in mitochondria and chloroplasts in eukaryotes (Porankiewicz *et al*, 1999). The *clpP* gene in *E. coli* encodes a polypeptide of 207 amino acids, of which 14 residues are autolytically cleaved to yield the mature protein of 21.5 kDa (Maurizi *et al*, 1990a). The active enzyme is a tetradecameric structure composed of two heptameric rings that form a hollow cylinder with 14 proteolytic sites compartmentalized within its central chamber (Flanagan *et al*, 1995; Shin *et al*, 1996; Wang *et al*, 1997). Substrates enter into the chamber through two axial openings with a diameter of about 10 Å, which limits the size of polypeptides degraded. By itself, *E. coli* ClpP is able to rapidly hydrolyse only oligopeptides, but not large globular proteins. The degradation of large proteins requires the presence of an AAA ATPase complex, such as ClpA or ClpX in *E. coli* or ClpC in other species (Kress *et al*, 2009). These hexameric structures associate with both ends of ClpP to form the active 4-ring ATP-dependent protease (Maurizi, 1991; Maurizi *et al*, 1998; Kim *et al*, 2001). These ATPases bind selectively certain protein substrates, unfold them, and translocate the linearized polypeptides into the ClpP proteolytic chamber for degradation (Hoskins *et al*, 1998; Ortega *et al*, 2000; Ishikawa *et al*, 2001; Reid *et al*, 2001). In addition to substrate recognition, the human mitochondrial ClpX complex promotes the assembly of the ClpP complex into an active form (Kang *et al*, 2005).

Most organisms possess a single *clpP* gene, while some microorganisms (e.g., *Streptomyces*, *Actinomycetes*, and *Cyanobacteria*) and plants (e.g., *Arabidopsis thaliana*) have two or more *clpP*s (Porankiewicz *et al*, 1999; Viala *et al*, 2000; Peltier *et al*, 2001, 2004; Schelin *et al*, 2002; Viala and Mazodier, 2002; Butler *et al*, 2006; Sjogren *et al*, 2006; Stanne *et al*, 2007; Andersson *et al*, 2009). The functional significance of these multiple species is unclear. Mtb contains

*Corresponding author. Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA. Tel.: +1 617 432 1854; Fax: +1 617 232 0173; E-mail: alfred_goldberg@hms.harvard.edu
[3]These authors contributed equally to this work

Received: 14 September 2011; accepted: 22 December 2011

two *clpP* genes, *clpP1* and *clpP2*, both of which are essential for viability (Sassetti *et al*, 2003) and infectivity, as shown in Raju *et al* (2011). Although both appear to encode serine proteases, prior attempts (Ingvarsson *et al*, 2007; Benaroudj *et al*, 2011) to express and characterize Mtb ClpP1 and ClpP2 in *E. coli* yielded complexes that lacked proteolytic activity, as did our initial attempts to express ClpP1 and ClpP2 in *E. coli*. We hypothesized that those attempts failed because they were based on the assumption that ClpP1 and ClpP2 are distinct enzymes, while in fact, the active enzyme *in vivo* is a mixed complex.

Here, we demonstrate that ClpP1 and ClpP2, when overproduced independently, form tetradecameric complexes that lack any proteolytic activity. However, when these complexes are mixed together in the presence of certain small activating molecules, which we accidentally discovered, these tetradecamers dissociate into heptameric rings, which then re-associate into a mixed tetradecameric complex that is capable of degrading model peptides as well as some unstructured proteins. These low molecular weight activators clearly represent a novel form of enzyme regulation and stimulate ClpP1P2 activity in a very different manner from the regulatory ATPase complex, ClpC1, which we show enhances specifically the degradation of proteins. Thus, ClpP1P2 differs markedly from other members of the ClpP family and has a number of highly unusual structural, enzymatic, and regulatory properties. These unique qualities of ClpP1P2, taken together with its essential role during infection, make it an attractive target for drug development.

Results

Isolation of processed ClpP1 and ClpP2

Mtb *clpP1* or *clpP2* genes were expressed as C-terminal fusions with 6 × His and/or Myc tags under the control of a tetracycline-inducible promoter. Since previous efforts and our initial attempts to produce active ClpP1 and ClpP2 in *E. coli* were unsuccessful (Ingvarsson *et al*, 2007; Benaroudj *et al*, 2011), we attempted to separately express ClpP1 and ClpP2 under conditions resembling those in Mtb by using the closely related non-pathogenic species *M. smegmatis*. Purification on an Ni-NTA agarose column yielded large amounts of nearly pure proteins, each with an apparent molecular weight of ~22 kDa (Figure 1A). When ClpP1 and ClpP2 were subjected to gel filtration on an S-300 Sephacryl column, both were eluted as single homogenous peaks with a molecular mass of about 300 kDa (Figure 3A, top panel). Thus, both ClpP1 and ClpP2 had the same elution profile as *E. coli* ClpP and appeared to be 14-subunit 2-ring complexes.

The ClpP1 and ClpP2 bands from the SDS–PAGE were digested by trypsin and chymotrypsin and analysed by MS/MS. Eighty-three peptides were identified for ClpP1 (92% coverage by amino acids) and 70 peptides for ClpP2 (94% coverage). Although mass spectrometry thus demonstrated nearly all the expected peptides, N-terminal sequencing indicated that ~70% of both proteins were N-terminally processed with major cleavage sites at Asp^{6}-Met^{7} for ClpP1 and Ala^{12}-Arg^{13} for ClpP2 (Figure 1B). (In addition, minor cleavages were also detected at Thr^{5}-Asp^{6} and Met^{7}-Arg^{8} for ClpP1 and Arg^{13}-Tyr^{14} for ClpP2.) It is noteworthy that the extent of this processing varied in different preparations and correlated with their ability to support enzymatic activity. Thus, N-terminal processing of both gene products appears important for the formation of the active enzyme. Moreover, when full-length mutant forms of ClpP1 and ClpP2, which lacked enzymatic activity (see below), were expressed in *M. smegmatis*, a much smaller fraction of N-terminally processed forms could be detected. Therefore, it is likely that the proteolytic processing of mycobacterial ClpPs occurs primarily through an autocatalytic mechanism (possibly involving collaboration with the *M. smegmatis* enzymes). Accordingly, ClpP1 is cleaved after Asp (Figure 1B), which as shown below, is one of the preferred sites for Mtb ClpP (see below, Table I).

In subsequent studies, we therefore expressed the constructs corresponding to the processed versions directly and obtained more homogenous preparations with higher activities. It is noteworthy that these shorter forms, which do not require N-terminal processing, could also be efficiently produced in *E. coli*.

ClpP1 and ClpP2 form a mixed ClpP1P2 protease that requires certain short peptides for activation

Neither ClpP1 nor ClpP2 alone had peptidase activity (Figure 1C), although both formed tetradecameric structures characteristic of the ClpP family. Because both genes are essential (Sassetti *et al*, 2003; Ollinger *et al*, 2012; Raju *et al*, 2011), we hypothesized that ClpP1 and ClpP2 are not two distinct enzymes, but instead associate to form a novel, mixed proteolytic complex. To test this possibility, we first attempted to co-express Mtb ClpP1 and ClpP2 in *M. smegmatis*. The two proteins associated *in vivo* since they could be co-immunoprecipitated from the cell extract (Raju *et al*, 2011). However, due to wide variations in the levels of ClpP1 and ClpP2 expression, the ratios between the co-purified ClpP1 and ClpP2 varied markedly in different preparations, and this heterogeneity prevented rigorous study of the active complex. Therefore, we expressed them separately and attempted to reconstitute such a mixed complex from pure components. In fact, mixing pure ClpP1 and ClpP2 together in high concentrations (up to 0.5 mg/ml) resulted in the appearance of very low peptidase activity against the fluorogenic substrate of *E. coli* ClpP, Suc-Leu-Tyr-AMC.

During attempts to identify transition state-specific inhibitors of this low activity, we accidentally made the surprising, but very valuable discovery that a group of N-blocked peptide aldehydes that were substrate analogues not only did not inhibit, but actually stimulated this activity over 1000-fold. A similar dramatic activation was even found with certain related blocked peptides. For example, as shown in Figure 1C, a mixture of ClpP1 and ClpP2 was inactive in hydrolysing the Z-Gly-Gly-Leu-AMC or the quenched fluorescent substrate Mca-GHQQYKMK-Dpa(Dnp)-amide, but in the presence of the activating peptide Z-Leu-Leu, both substrates and the unfolded protein, casein, were efficiently cleaved (Figure 1C and D). The activating peptides and peptide aldehydes only induced peptidase activity if both ClpP1 and ClpP2 were present together. It is noteworthy that at 37°C (under our standard assay conditions) the activation occurred without any noticeable delay after the addition of the peptide activator. Also, the activator had to be continually present for enzymatic activity. When the activator was removed by gel filtration or if its concentration was reduced by dilution, activity was lost, but it could be regained

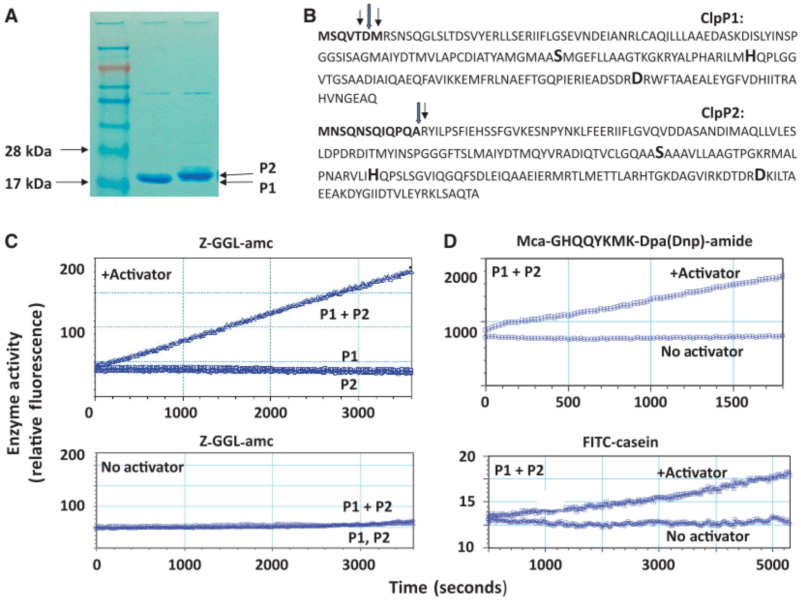


Figure 1 Purification of processed but inactive ClpP1 and ClpP2 and reconstitution of the active ClpP1P2 complex. (**A**) Coomassie staining of ClpP1 (2.3 μg) and ClpP2 (2.7 μg) after purification. (**B**) Sequences of ClpP1 and ClpP2 proteins expressed in *M. smegmatis*. Arrows indicate the sites of proteolytic processing determined by Mass Spectrometry and N-terminal sequencing of ClpP1 and ClpP2 purified as in (**A**). (**C**) ClpP1P2 (2.1 μg) possesses peptidase activity but only in the presence of activating dipeptide Z-Leu-Leu. ClpP1 (1.8 μg) or ClpP2 (2.4 μg) alone did not show any activity with or without the activator. Enzymatic activity was measured fluorometrically using Z-Gly-Gly-Leu-AMC as a substrate. (**D**) Activator also stimulates degradation of longer peptides (Mca-KKPTPIQLN-Dpa(Dnp)-amide) and proteins (FITC-casein) by ClpP1P2 (0.36 and 2.9 μg, respectively).

fully upon restoration of activator to its prior concentration (Figure 2C).

The strongest stimulation against Z-Gly-Gly-Leu-AMC, as well as other substrates, was found with Z-Leu-leucinal (Figure 2), but the longer aldehyde Z-Leu-Leu-leucinal was significantly less active. Several other hydrophobic dipeptide aldehydes (e.g., Z-Val-phenylalaninal), acidic peptide aldehydes (e.g., Z-Pro-Nle-aspartal) and alkyl aldehydes did not show any stimulatory capacity. The effective peptide aldehydes presumably should bind to at least some of the enzymes' 14 active sites. However, the related peptide Z-Leu-Leu and its alcohol derivative Z-Leu-leucinol (which presumably should not bind strongly to the active sites) could also activate ClpP1P2, although only at much higher concentrations than the corresponding aldehydes. A much smaller stimulation was observed with blocked peptides Z-Leu, Z-Gly-Leu and Z-Gly-Leu-Leu (Figure 2A).

The concentration dependence for activation, by Z-Leu-leucinal (Kd = 0.24 mM) and Z-Leu-Leu (Kd = 2.2 mM), revealed a highly cooperative mechanism with a Hill coefficient of 5–7 (Figure 2B). Thus, multiple molecules probably bind to ClpP1P2 to stimulate its activity. Though substrate analogues, these activators are not cleaved, since upon incubation with ClpP1P2, no new amino groups could be detected using the sensitive fluorescamine assay. It is noteworthy that although the aldehyde had a higher affinity, at high concentrations, Z-Leu-Leu caused a greater activation than Z-Leu-leucinal (Figure 2B). Also because peptides are more stable and much less expensive than the corresponding aldehydes, in subsequent studies, we routinely induce Mtb ClpP1P2 activity using Z-Leu-Leu (subsequently referred to as the 'activator').

Activation involves dissociation of ClpP1 and ClpP2 tetradecamers and formation of 2-ring ClpP1P2 complex

Because the activators stimulate only ClpP1 and ClpP2 together (but not pure ClpP1 or ClpP2; Figure 1C and D), they probably activate by promoting the formation of a new mixed ClpP1P2 complex. We therefore examined how the presence of an activator affects the sizes of these different complexes. Upon size-exclusion chromatography, a mixture of ClpP1 and ClpP2 behaved as tetradecamers exactly like pure Mtb ClpP1 or ClpP2 and *E. coli* ClpP (Figure 3A, upper panel). However, when the activator Z-Leu-Leu was present (Figure 3A, lower panel), both ClpP1 and ClpP2 peaks were eluted as a single lower molecular weight peak, resembling γ-globulin (150 kDa) in size. Thus, the tetradecameric (presumably 2-ring) complexes composed of a single subunit type dissociated into heptamers. However, in the presence of the activator, the ClpP1/ClpP2 mixture was eluted as a 300-kDa peak that coincided with the peptidase activity and corresponded in size to ClpP tetradecamers (Figure 3A, lower panel). The ClpP1P2 complexes were isolated from the peak using Ni-NTA (by His-tagged ClpP2) or anti-Myc (by Myc-tagged

Table I Mtb ClpP1P2 preferentially hydrolyses peptides with hydrophobic and acidic residues in P1 position

Peptide substrate	Relative rates of hydrolysis (%)
Hydrophobic P1 residue	
Z-Gly-Gly-Leu-amc	100.0
Suc-Leu-Leu-Val-Tyr-amc	0.8
Suc-Leu-Tyr-amc	11.5
Z-Leu-Leu-Leu-amc	0.15
Z-Leu-Leu-amc	4.7
Z-Ala-Ala-Ala-amc	3.6
Suc-Ala-Leu-Pro-Phe-amc	0.12
Suc-Ala-Ala-Pro-Ala-amc	0.08
Ala-Ala-Phe-amc	87.0
Suc-Ala-Ala-Phe-amc	42.0
Acidic P1 residue	
Ac-Nle-Pro-Nle-Asp-amc	36.0
Z-Leu-Leu-Glu-amc	0.35
Basic P1 residue	
Z-Leu-Leu-Arg-amc	0.25
Z-Gly-Gly-Arg-amc	0.55
Z-Phe-Val-Arg-amc	0.18
Aminopeptidase substrates	
Ala-amc	0.78
Leu-amc	0.36
Phe-amc	0.26
Asp-amc	0.1

Peptidase activity of ClpP1P2 (1.8 μg) was measured with a variety of substrates (0.1 mM) and compared with activity obtained with Z-Gly-Gly-Leu-AMC (100%).

ClpP1) columns, and the presence of both proteins in resin-bound material was confirmed by MS (see Materials and methods).

Thus, the activating peptide causes the dissociation of ClpP1 and ClpP2 tetradecamers into heptamers and favours their subsequent association to form the active tetradecameric ClpP1P2 complex. By contrast, no changes in elution pattern were observed when *E. coli* ClpP was incubated with this activator.

Conformational changes accompanying formation of ClpP1P2 complex

The dissociation and re-association of multimeric ClpP1 and ClpP2 rings must involve activation-induced major changes in subunit conformation. Because ClpP1 (but not ClpP2) contains a Trp residue, we can use it to monitor conformational changes that may accompany the formation of an active ClpP1P2 complex from inactive ClpP1 and ClpP2 ones. Although no spectral changes were observed with dissociation of the ClpP1 tetradecamer upon addition of the activator, the formation of the active ClpP1P2 complex appears to involve changes in ClpP1's conformation, because the fluorescence of Trp174 in ClpP1 shifted its maximal fluorescence from 345 in pure ClpP1 to 338 nm. (Figure 3B). Thus, the interaction between ClpP1 and ClpP2 subunits leading to activation is associated with changes in the subunits' conformation. It is noteworthy that similar changes in Trp174 fluorescence occurred when active-site mutants of ClpP1 and ClpP2 that lack enzymatic activity (see below) were mixed in the presence of the activator. Thus, enzymatic activity of both ClpPs is not necessary for their dissociation—re-association and the major structural changes associated with this activation process.

To confirm that such a mixed ClpP1P2 complex actually exists *in vivo*, we tested whether endogenous ClpP1 and ClpP2 associate in wild-type *M. smegmatis*. As described in the related manuscript (Raju *et al*, 2011), we employed mycobacterial recombineering to add a C-myc tag to the C-terminus of genomic ClpP2. The C-myc-tagged ClpP2 was isolated together with associated proteins using an anti-myc resin, and the material eluted with the Myc peptide was resolved by SDS–PAGE. Bands corresponding by size to ClpP2 and ClpP1 were analysed by Mass Spectrometry, and the presence of both subunits was confirmed, indicating that mixed ClpP1P2 complexes are present in mycobacteria.

Mtb ClpP1P2 is composed of one ClpP1 and one ClpP2 heptameric ring

To determine the subunit composition of this ClpP1P2 complex, we varied the relative concentrations of ClpP1 and ClpP2 in the presence of an activator (Figure 4A). Upon increasing the amount of ClpP1 with a constant amount of ClpP2, peptidase activity gradually increased and reached its maximum when these components were present in close to equimolar amounts. Conversely, when ClpP1 content was held constant and the amount of ClpP2 increased, maximal activity was also obtained with equimolar concentrations. In different experiments using different ClpP1 and ClpP2 preparations, the optimal ClpP1/ClpP2 molar ratio ranged from 0.82 to 1.15. Thus, the active complex contains equal numbers of ClpP1 and ClpP2 subunits.

These findings and the rapidity of activation together strongly suggest that the active enzyme is composed of one ClpP1 and one ClpP2 ring. However, it is also possible that each heptameric ring contains a mixture of ClpP1 and ClpP2 subunits, as has been found for the cyanobacterium *Synechococcus* ClpP complexes (Stanne *et al*, 2007; Andersson *et al*, 2009). To determine the composition of the rings, we crosslinked the neighbouring subunits in the active ClpP1P2 tetradecamer with glutaraldehyde in the presence of the activator (Figure 4B). After a 0.5 h of incubation, seven distinct crosslinked bands were evident on SDS–PAGE corresponding to 1-, 2-, 3- 4-, 5-, 6-, and 7-mers. As expected, the larger crosslinked structures were the least abundant. After an overnight incubation, when crosslinking went to completion, all seven subunits, presumably comprising the rings, were crosslinked together, but still no band was observed with a molecular mass higher than that of a 7-mer. Thus, apparently, no crosslinking occurred between the two rings (which presumably requires a crosslinker with a longer spacer arm than glutaraldehyde). Analysis by mass spectrometry indicated that the crosslinked heptamers were composed only of ClpP1 or ClpP2 subunits, respectively, and no peptides corresponding to ClpP1–ClpP2 crosslinked were found. Thus, each heptameric ring in the Mtb ClpP1P2 protease is homogenous in composition.

Cleavage specificity of Mtb ClpP1P2

To define the substrate preference of the ClpP1P2 active sites, we tested a variety of synthetic fluorescent peptides with hydrophobic, acidic, or basic residues in the P1 position (Table I). The best substrate was Z-Gly-Gly-Leu-AMC, while Suc-Ala-Ala-Phe-AMC and Ala-Ala-Phe-AMC also were readily cleaved. The failure of ClpP1P2 to degrade rapidly the widely used proteasome substrate Suc-Leu-Val-Tyr-AMC

A Certain short peptide aldehydes and peptides dramatically activate Mtb ClpP1P2

Compound	Relative activity (%)
Peptide aldehydes (0.5 mM)	
Z-Leu-Norleucinal	100.0
Z-Leu-Leucinal	77.6
Z-Val-Phenilalaninal	1.9
Z-Phe-Tyrosinal	1.3
Z-Pro-Nlee-Asparaginal	1.2
Z-Ala-Pro-Nle-Asparaginal	0.8
Peptide derivatives (1 mM)	
Z-Leu-Leucinal	100.0
Z-Leu-Leu	18.5
Leu-Leu	0.1
Z-Leu	5.3
Z-Leu-Leucinol	17.6
Z-Leucinol	2.4
Z-Gly-Leu	4.8
Z-Gly-Gly-Leu	3.7

C Activation of ClpP1P2 by Z-Leu-Leu is readily reversible

Active enzyme, diluted (200x)	Relative activity (%)
With activator	100.0
Without activator	1.2
Without activator, then activator re-added	96.0

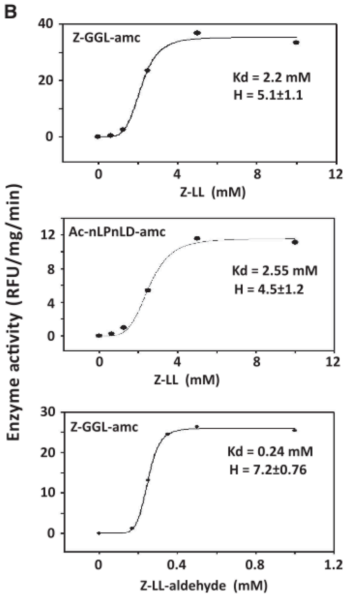


Figure 2 Certain short peptides and peptide aldehydes dramatically activate ClpP1P2 by binding to multiple sites. (**A**) Effects of various short peptide aldehydes and related peptides or peptide alcohols on activity of ClpP1P2 (1.3 μg). Peptidase activity was measured with Z-Gly-Gly-Leu-AMC. The specific activity with Z-Leu-norleucinal was 4.25 μmole/mg/min, which was taken as 100%. (**B**) Determination of Hill coefficient of Z-Leu-Leu and Z-Leu-leucinal in the hydrolysis of Z-Gly-Gly-Leu-AMC (0.1 mM) and Ac-Nle-Pro-Nle-Asp-amc (0.1 mM) by ClpP1P2 (2.4 μg). (**C**) Activation of ClpP1P2 by Z-Leu-Leu is readily reversible. Enzyme (1.8 μg) was incubated in the presence of activator Z-Leu-Leu (5 mM) and then diluted 200-fold in the buffer with and without the activator. Enzymatic activity was assayed with Z-Gly-Leu-Leu-AMC (0.1 mM). Re-addition of the activator restored the ClpP1P2 activity completely.

indicates major differences from enzymes in the mammalian cytosol. It is noteworthy that Z-Leu-Leu-AMC, the fluorescent peptide corresponding to the peptide activator employed routinely Z-Leu-Leu, was a poor substrate for the enzyme (Table I), and conversely the peptides corresponding to the best substrates, Z-Gly-Gly-Leu or Z-Gly-Leu, were poor as activators (Figure 2A).

In addition to hydrophobic peptides, ClpP1P2 also efficiently hydrolyses a peptide with acidic residues in the P1 position, Ac-Nle-Pro-Nle-Asp-AMC. (We also found that this substrate is degraded by *E. coli* ClpP, which had been reported to cleave after aspartate residues in model polypeptides; Thompson and Maurizi, 1994). However, Mtb ClpP1P2 did not hydrolyse Z-Leu-Leu-Glu-AMC or peptides with basic P1 residue and was also inactive against a variety of unblocked amino acid-AMC substrates used to assay aminopeptidases (Table I). ClpP1P2 also could cleave a variety of longer quenched fluorescent peptides (e.g., Mca-GNTQFKRR-Dpa(Dnp)-amide, Mca-GHQQYAMK-Dpa(Dnp)-amide, Mca-GNQQYKMK-Dpa(Dnp)-amide and Mca-KKPTPIQLN-Dpa(Dnp)-amide), and could degrade slowly the largely unstructured protein FITC-casein, provided an activator was present (Figure 1D).

Though ClpP1 or ClpP2 alone lack enzymatic activity, their catalytic triads are formed

The sequences of both ClpP1 and ClpP2 appear to contain a Ser/His/Asp catalytic triad characteristic of serine proteases (Figure 1B). Accordingly, Mtb ClpP1P2 was sensitive to most standard inhibitors of serine proteases (Figure 5A), including agents that react with the active-site serine (dichloroisocoumarin, Powers and Kam, 1994 and biotinylated derivative of fluoroethoxyphosphinyl (FP-biotin), Liu *et al*, 1999), and peptide chloromethyl ketones (Szyk and Maurizi, 2006), which modify the catalytic histidine. By contrast, standard inhibitors of metalloproteases and cysteine proteases had no effect. Interestingly, the hydrolysis of both hydrophobic and acidic substrates was inhibited to similar extents by the peptide chloromethyl ketones, Z-LY-CMK or AAF-CMK.

To learn whether both gene products are enzymatically active in the complex, we incubated ClpP1, ClpP2, and the ClpP1/ClpP2 mixture with the biotinylated covalent modifier of active-site serines. As shown in Figure 5B, both ClpP1 and ClpP2 subunits were covalently modified even in the enzymatically inactive ClpP1 or ClpP2 complexes. Thus, the catalytic triad appears functional in these complexes despite their lack of enzymatic activity. To test the possibility of a non-specific binding of biotinylated modifier, the mutant forms of ClpP1 or ClpP2 with active-site Ser substituted for Ala were incubated with FP-biotin. As shown in Figure 5B, no incorporation of the modifier in the mutant proteins occurred, thus confirming its specific reaction with the active sites.

We anticipated that the activator would enhance the modification of the active-site serines by promoting active-site formation. Surprisingly, the presence of the activator did

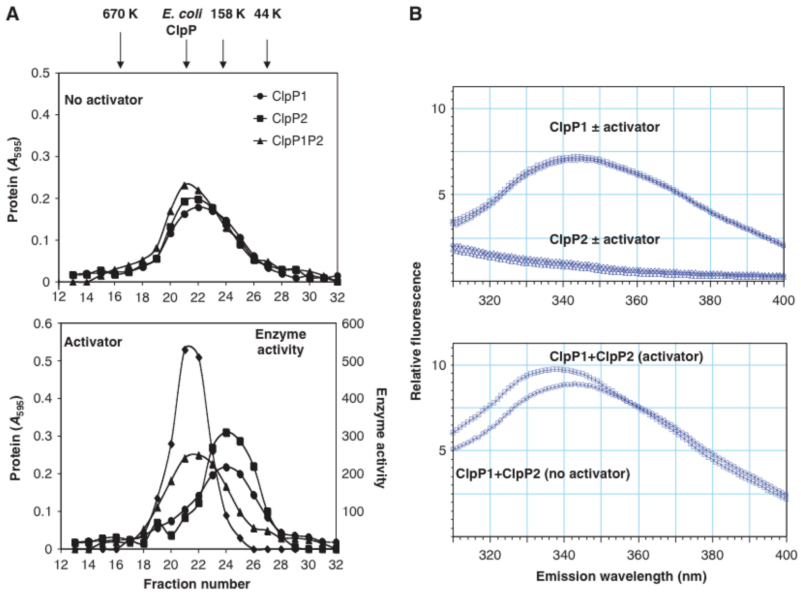


Figure 3 With the activator present, ClpP1 and ClpP2 tetradecamers dissociate into heptamers but ClpP1/ClpP2 mixture forms an enzymatically active tetradecamer. (**A**) Size-exclusion chromatography of ClpP1 (0.28 mg), ClpP2 (0.34 mg), and ClpP1/ClpP2 mixture (0.31 mg) was carried out using Sephacryl S300 column in the absence (upper panel) and presence (lower panel) of activator Z-Leu-Leu, which was also present in the running buffer. Samples (0.2 ml) were applied to the column and 0.5 ml fractions were collected and assayed for protein content and activity with Z-Gly-Gly-Leu-AMC. The column was calibrated with γ-thyroglobulin (670 K), γ-globulin (158 K), ovalbumin (44 K), and *E. coli* ClpP (300 K). (**B**) The change in fluorescence emission spectrum of ClpP1/ClpP2 mixture (12.8 μg in 100 μl) upon addition of activator indicates a conformational change due to complex formation between ClpP1 and ClpP2. The addition of activator to ClpP1/ClpP2 mixture shifted the peak of ClpP1 Trp174 emission from 345 to 338 nm (lower panel), while no change was observed with ClpP1 alone (upper panel).

not stimulate the modification of either homogeneous ClpP1 or ClpP2 or the mixed complex. In fact, the activator even reduced slightly (but reproducibly) this reaction, as also occurred in the presence of a substrate. Similar inhibition was observed in the presence of a substrate. Thus some activator molecules, which are structurally related to peptide substrates, appear to bind to the active sites. In any case, because active-site labelling occurred with ClpP1 and ClpP2 alone, these results prove that activation is not through formation of the catalytic triad, and instead that formation of the mixed tetradecamer probably enables the substrate to access and bind to the previously latent active sites.

ClpP1 and ClpP2 have distinct cleavage specificities

To estimate how ClpP1 and ClpP2 subunits contribute to enzymatic activity of the complex, pure ClpP1 or ClpP2 was inactivated by pretreatment with dichloroisocoumarin at concentrations that completely inhibit ClpP1P2. When the covalently inactivated ClpP1 or ClpP2 was incubated with its normal counterpart plus activator, the hydrolysis of hydrophobic and acidic peptide substrates, as well as casein, was significantly less than with untreated subunits (Figure 5C). Thus, both types of subunits appear to function enzymatically and contribute to the activity of the complex. However, inactivation of ClpP1 caused a much greater loss of these activities than did inactivation of ClpP2, especially with the hydrophobic peptide substrate (Figure 5C).

To confirm these different roles of each subunit, we reconstituted enzymatic complexes using wtClpP1 and an active-site mutant ClpP2 (Ser^{110} to Ala) or with wtClpP2 and active-site mutant ClpP1 (Ser^{98} to Ala). The complexes containing only one type of active subunits showed lower peptidase activity than the wild-type enzyme against both hydrophobic and acidic peptide substrates, and casein (Figure 5C). As was found upon derivatization with isocoumarin, the lack of functional ClpP1 caused a greater loss of activity against these various substrates than did the loss of ClpP2, particularly with the hydrophobic substrate. These data indicate that ClpP1 and ClpP2 active sites have different substrate preferences and suggest that ClpP1 sites are more important than ClpP2's in cleaving the most abundant bonds in proteins.

The ClpC1 ATPase complex stimulates protein degradation by ClpP1P2 but only in the presence of both an activator and ATP

The mechanism for the dramatic ClpP1P2 activation by small peptides uncovered here was surprising and unprecedented.

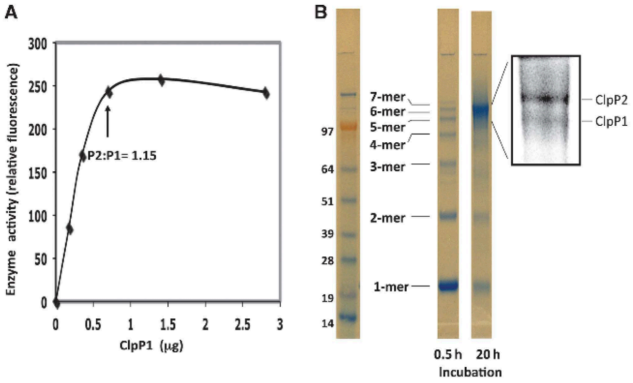


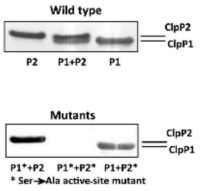
Figure 4 ClpP1P2 has maximal activity when equimolar amounts of ClpP1 and ClpP2 are present (**A**) and is composed of heptameric rings containing only ClpP1 or ClpP2 subunits (**B**). (**A**) Activity of ClpP1P2 at different ClpP1:ClpP2 ratios. Constant amounts of ClpP2 (0.85 μg) were mixed with increasing amounts of ClpP1, and Z-Gly-Gly-Leu-AMC hydrolysis was measured in the presence of activator. (**B**) Crosslinking of ClpP1P2 subunits by glutaraldehyde. After 0.5 and 20 h incubation at room temperature of ClpP1P2 (12 μg) with 0.125% glutaraldehyde, the reaction mixture was analysed by SDS–PAGE, followed by Mass Spectrometry. High molecular weight bands corresponding to seven crosslinked subunits were found to contain exclusively ClpP1 or ClpP2 peptides, indicating that each ring contains seven identical ClpP1 or ClpP2 subunits. In the insert, one tenth of the crosslinked material used in Figure 4B was resolved by SDS–PAGE and stained by Coomassie Blue. The gel was scanned, and the image enlarged using Photoshop.

A Mtb ClpP1P2 has inhibitor specificity characteristic of serine endopeptidases

Compound	% Inhibition
3,4-Dichloroisocoumatin (0.1 mM)	100
Z-Leu-Tyr-CMK (0.1 mM)	78
Ala-Ala-Phe-CMK (0.1 mM)	69
FP-Biotin* (0.1 mM)	45
PMSF (2 mM)	15
Leucinethiol (0.05 mM)	5
Bestatin (0.05 mM)	3
Iodoacetic acid (1 mM)	0
1,10-Phenanthroline (0.25 mM)	12
EDTA (1 mM)	5

*10-(fluoroethoxyphosphinyl)-N-(biotinamidopentyl)decanamide
CMK- chloromethylketone

B Biotinylated active-site modifier (FP-Biotin) binds to WT ClpP1 and ClpP2 subunits but not to active-site mutants



C Inactivation of ClpP1 or ClpP2 indicates that they possess different substrate preferences

Effects of inactivation of ClpP1 or ClpP2 on ClpP1P2 function					
Enzyme	WT	P2 inactivated		P1 inactivated	
ClpP1	WT	WT	WT	DCL treated	Mutated
ClpP2	WT	DCL treated	Mutated	WT	WT
Substrates		Relative activity (%)		Relative activity (%)	
Z-GGL-amc	100	78	67	3	5
Ac-nLPnLD-amc	100	36	45	21	27
FITC-casein	100	Not available	82	Not available	30

Figure 5 Both ClpP1 and ClpP2 contain functional active sites, but ClpP1's active sites are more important than ClpP2's in ClpP1P2 activity. (**A**) ClpP1P2 protease has the inhibitor sensitivity characteristic of serine proteases. Peptidase activity against Z-Gly-Gly-Leu-AMC (0.1 mM) was measured after 30 min preincubation of the enzyme (1.2 μg) with or without inhibitors at room temperature. The error ranges of the inhibition did not exceed 5%. (**B**) To learn whether the active-site charge relay system is functional in the absence of activator, 1.2 μg of wild-type or active-site Ser-Ala mutants of ClpP1, ClpP2, or a ClpP1/ClpP2 mixtures were incubated for 30 min with a biotinylated derivative of an active-site inhibitor fluoroethoxyphosphinyl. The binding of the inhibitor was determined by SDS–PAGE followed by western blot analysis with an anti-biotin antibody. (**C**) To determine how ClpP1 and ClpP2 contribute to the enzymatic activity of ClpP1P2, either ClpP1 or ClpP2 was inactivated separately by pretreatment with dichloroisocoumarin (0.1 mM) or by an active-site mutation (active site Ser to Ala). The inactivated ClpP1 or ClpP2 was then mixed with its normal (untreated WT) counterpart in the presence of activator Z-Leu-Leu, and hydrolysis of hydrophobic and acidic peptide substrates and casein was measured. FITC-casein was not used to measure proteinase activity of DCL-treated enzymes due to the overlapping of the fluorescence spectra of FITC and DCL. Figure source data can be found in Supplementary data.

One attractive possible mechanism would be that the conformational changes and complex formation induced by the activator resemble those changes caused by the binding of the Mtb regulatory ATPase complex, ClpC1 or ClpX, since in some ATP-dependent proteases (e.g., human mitochondrial ClpP), no peptidase activity was demonstrated in the absence of nucleotides and the regulatory ATPase. To address the possibility that the peptide activator mimicked the regulatory ATPases, we cloned the ClpC1 ATPase from Mtb, expressed it in *M. smegmatis*, and tested whether it can stimulate ClpP1P2 activity. As expected, the resulting complex was of high molecular weight and had ATPase activity. (The isolation and characterization of Mtb ClpC1 will be reported in detail in a separate manuscript.) Unlike the dipeptide activator, addition of ClpC1 did not increase the hydrolysis of any of various fluorescent peptides assayed, nor did it enhance the effect of the activator on their degradation (Figure 6A). However, when the ATPase was added in the presence of the activator, it markedly increased the degradation of the model protein, FITC-casein (Figure 6A). This stimulation of protein degradation by ClpP1P2 was only observed when both the activator and ATP were present (Figure 6C). Thus, ClpC1 and the activator must increase proteolysis by quite different, but clearly additive mechanisms.

It is noteworthy that ClpC1 did not stimulate proteolysis by ClpP1 or ClpP2 alone (Figure 6B). Therefore, it is very likely that ClpC1 *in vivo* also functions only with the mixed ClpP1P2 complex and requires an additional factor resembling the activator for protein degradation. The peptide activator did not influence ATP hydrolysis by ClpC1. By contrast, the protein substrate casein stimulated ClpC1's ATPase activity two-fold (Figure 6D) in a similar fashion to the activation by substrates of the homologous *E. coli* AAA ATPases, Lon (Waxman and Goldberg, 1982), ClpA (Hwang *et al*, 1988; Thompson and Maurizi, 1994), and HslU (Seol *et al*, 1997). Thus, the peptide activator is necessary only for ClpP1P2 assembly, while ClpC1 binds casein directly and facilitates its degradation by ClpP1P2.

Discussion

Mtb ClpP1P2 is a novel enzyme complex in multiple respects

Several observations led us to hypothesize that ClpP1 and ClpP2 function together in a single complex. (1) Our initial attempts and those of others (Ingvarsson *et al*, 2007; Benaroudj *et al*, 2011) to isolate active ClpP1 or ClpP2 from Mtb were unsuccessful, even though they formed tetradecameric complexes similar in size to other ClpP family members, as shown in Figure 3 and by others (Ingvarsson *et al*, 2007). (2) In our related paper (Raju *et al*, 2011), we showed that both *clpP1* and *clpP2* genes are essential for growth and infectivity of Mtb, and thus, they cannot compensate for the loss of the other. (3) When ClpP1 and ClpP2 were co-expressed in *M. smegmatis*, they could be co-immunoprecipitated (Raju *et al*, 2011). We then showed that neither purified ClpP1 nor ClpP2 alone is enzymatically active, but if they were both present, together with an activating peptide or peptide derivative, then mixed complexes were formed that showed robust proteolytic activity. Furthermore, only when the mixed ClpP1P2 complex was formed in the presence of the activator, casein degradation was stimulated in an ATP-dependent manner by ClpC1 ATPase. Since this latter process resembles the conditions for protein degradation *in vivo* and since ClpC1 is also essential for viability, it seems very likely that the ClpP1P2 is the functional protease *in vivo*.

One initial approach that enabled us to obtain active ClpP1P2 was the use of an unusual expression system.

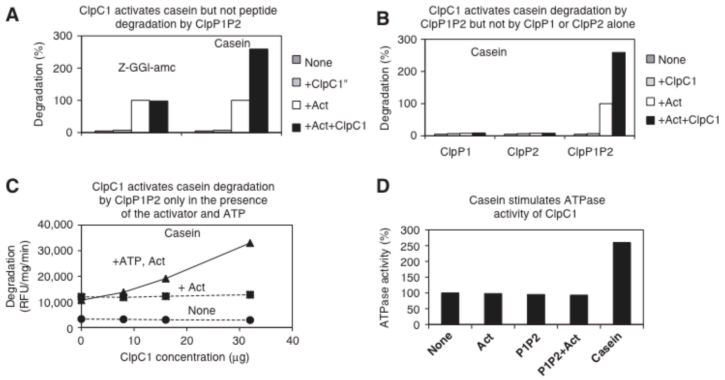


Figure 6 ClpC1 activates degradation of casein, but not peptides, by ClpP1P2 only in the presence of activator and ATP. (**A**) ClpC1 activates casein but not peptide degradation by ClpP1P2. ClpP1P2 (2.5 μg) and ClpC1 (32 μg) were mixed in 100 μl of reaction buffer containing 2 mM ATP and 8 mM $MgCl_2$ (see Materials and methods) in the presence or absence of the peptide activator. Enzymatic activity was measured fluorometrically using Z-Gly-Gly-Leu-AMC or FITC-casein as substrates. The rate of degradation of ClpP1P2 in the presence of the activator was taken as 100%. (**B**) ClpC1 activates casein degradation by ClpP1P2 but not by ClpP1 or ClpP2 alone. Degradation of FITC-casein by ClpP1 (2.5 μg), ClpP2 (2.9 μg), or ClpP1P2 (2.7 μg) was measured as in (**A**). (**C**) The stimulation of FITC-casein degradation by ClpP1P2 requires ATP and the peptide activator. Degradation was measured as in (**A**) and (**B**) with different concentrations of ClpC1. (**D**) Casein stimulates ATPase activity of ClpC1. In all, 2 μg of ClpC1 was incubated with 5 μg FITC-casein, or with ClpP1P2 +/− activator, or with the activator alone, in 20 μl of reaction buffer for 30 min and the ATPase activity was measured by Malachite Green method (see Materials and methods).

Expressing the cloned Mtb ClpP1 and ClpP2 proteins in *E. coli* did not yield the mature, processed subunits, although some limited processing of Mtb ClpP1 and ClpP2 (by ClpP1) has been recently reported in *E. coli* (Benaroudj *et al*, 2011). In the present studies, we used the closely related mycobacterium, *M. smegmatis*, in which ClpP1 and ClpP2 underwent efficient N-terminal processing, which appears necessary for enzymatic activity. In fact, in the crystal structure of the inactive ClpP1 tetradecamer composed of unprocessed subunits, the distance between the active-site Ser and His residue was too large to support the formation of the active catalytic triad (Ingvarsson *et al*, 2007). Once we had identified the N-termini of the fully active ClpP1 and ClpP2, we directly expressed these shorter sequences, which yielded homogenous tetradecamers in *M. smegmatis* as well as in *E. coli.* Also, by using an *E. coli* mutant lacking endogenous ClpP, we could obtain WT or mutant ClpP1 and ClpP2 without contamination by endogenous ClpP subunits. It is noteworthy that these ClpP1 tetradecamers composed only of 'processed' subunits do contain a functional catalytic triad, although it is still unable to catalyse peptide hydrolysis. Such a lack of catalytic activity despite the presence of a functional catalytic triad has been previously shown for other pro-enzymes, such as trypsinogen, which though enzymatically inactive can react with various active-site titrants (Smith *et al*, 1992). Its lack of enzymatic activity is attributed to an inability of substrates to access the active sites, and presumably a similar explanation accounts for the inactivity of pure ClpP1 and ClpP2 (see below).

Recently, Mtb ClpP2 and ClpC1 were reported to catalyse the degradation of an endogenous Mtb protein Rse A (Barik *et al*, 2010). While degradation of that substrate perhaps may not require formation of a ClpP1P2 complex (unlike casein or the peptides studied here), several features of that study are difficult to reconcile with our findings. For example, they expressed Mtb ClpP proteins in *E. coli*, which we and others (Ingvarsson *et al*, 2007; Benaroudj *et al*, 2011) found to yield mostly non-processed inactive tetradecamers. Possibly, their use of *WT E. coli* for expression may have resulted in contamination by the highly active *E. coli* ClpP, as was reported by Benaroudj *et al* (2011). To avoid these problems, we expressed the mature Mtb proteins only in *E. coli* strain lacking ClpP.

Very little is known about ClpP proteases from organisms that contain multiple *clpP* genes. In *Streptomyces lividans*, five *clpP* genes are organized in two operons (*clpP1* and *clpP2; clpP3* and *clpP4*), and one is monocistronic (Viala and Mazodier, 2002). Both ClpP1 and ClpP2 are required for degradation of the transcriptional activator PopR, which suggests that they also form a single mixed complex (Viala and Mazodier, 2002). ClpP3 and ClpP4 can also function together in PopR degradation, and their coordinate regulation suggests that they also comprise a mixed complex (Viala and Mazodier, 2002). In plant organelles, the organization of the ClpP proteases is much more complex (Peltier *et al*, 2001, 2004); for example, tetradecameric ClpP complexes have been isolated from *Arabidopsis thaliana* that contain five different ClpP proteins and six different non-proteolytic ClpP homologues (ClpR) (Peltier *et al*, 2004). Although their composition and activities have not been studied, the different ClpP and ClpR proteins may be present in the same tetradecameric complex (Peltier *et al*, 2004). In fact, a novel form of ClpP has recently been characterized from the cyanobacterium *Synechococcus* that contains two identical heptameric rings, composed of three active ClpP3 and four inactive ClpR subunits (Andersson *et al*, 2009), which though inactive, are essential for the ClpC-dependent proteolytic activity.

Our crosslinking experiments demonstrate that Mtb ClpP1P2 heptameric rings are composed of seven identical subunits. Therefore, the active Mtb enzyme must be composed of one ClpP1 ring and one ClpP2 ring. Accordingly, optimal activity was obtained when ClpP1 and ClpP2 were present in a 1:1 molar ratio (Figure 4A). This association of the ClpP1 and ClpP2 rings with each other causes conformational changes that allow both complimentary rings to become enzymatically active. By monitoring the changes in tryptophan fluorescence in ClpP1 that accompany activation, we confirmed that the active ClpP1 conformation is achieved only in the presence of ClpP2 and a dipeptide activator. Interestingly, the ClpP1 and ClpP2 rings can activate each other, even if either or both were inactivated by mutation or derivatization of their active-site serines.

Although only a limited number of peptide substrates were screened, Mtb ClpP1P2 clearly has a rather broad substrate specificity. In its preference for large hydrophobic residues in the Pl position, ClpP1P2 resembles chymotrypsin; however, it also cleaves a peptide with an acidic residue in the P1 position that is a typical substrate for caspases and the caspase-like site on the proteasome (Kisselev *et al*, 2003, 2006). The active sites on ClpP1 and ClpP2 clearly differ in their cleavage specificity. ClpP1 clearly is predominant in the hydrolysis of casein and after hydrophobic residues (which are highly abundant in cell proteins). Its loss also reduces the rate of hydrolysis of acidic peptides but ClpP2 rings also contribute significantly to this activity. Further analysis of the hydrolysis of other substrates should be of value in the development of selective inhibitors.

Activation of ClpP1P2 by small molecules

The most unexpected and novel aspect of these findings is the discovery of small molecules that dramatically activate ClpP1P2 and enable it to degrade even unstructured polypeptides. *In vitro*, these agents were essential for both the appearance and the maintenance of enzyme activity. The most potent among these activators are short N-blocked peptide aldehydes, but the corresponding peptide alcohols and peptides also stimulate, though at higher concentrations. (Z-Leu-Leu was used routinely here because it is inexpensive and yields the largest maximal activation.) All these compounds markedly stimulate hydrolysis of all peptide substrates tested, as well as casein. In fact, without this surprising finding, the remaining observations on ClpP1P2 would not have been possible because its inherent activity is too low for most studies.

Even though ClpP1 and ClpP2 were enzymatically inactive by themselves, their active sites, even in the absence of the activator, could react with agents that covalently modify active-site serines or histidines, and did so as strongly as in the active enzyme (Figure 5B). Thus, in the absence of the activator, the catalytic triads appear to be functional in ClpP1 and ClpP2, unlike in the unprocessed ClpP1 (Ingvarsson *et al*, 2007). Presumably, these sites are unable to hydrolyze peptides in the absence of the complementary ring, because of a

failure of the substrate to enter the pure ClpP1 and ClpP2 complexes, as occurs with the latent form of the 20S proteasome, which is activated by a gating mechanism allowing substrate entry (Smith *et al*, 2007). Alternatively, formation of the mixed complex may involve structural rearrangements that enable catalysis.

The exact site where the activator binds to ClpP1P2 remains uncertain, and several possibilities exist. The structures of the activators closely resemble those of some hydrophobic substrates (Figure 2A; Table I), which suggest that the activators bind to the active sites. Accordingly, peptide aldehydes, which should bind tightly to active-site serine, were the strongest activators, while related peptides (which resemble products of substrate cleavage) are about 10 times less potent. Another observation supporting binding to the active sites was that the addition of activators instead of increasing enzyme interaction with the active-site titrant, actually decreased the extent of this modification. Because the structural changes that accompany ClpP1P2 activation (dissociation of tetradecamers into heptamers, formation of mixed complexes and changes in Trp fluorescence) were also induced by the activator in the inactive ClpP1 and ClpP2 active-site mutants, these activating dipeptides do not require interaction with the catalytic serines to induce dissociation—re-association.

Although agents that bind to the active sites should be competitive inhibitors if they bind to all the active sites, in the tetradecameric HslV protease complex from *E. coli*, Chung and coworkers have elegantly shown that inactivation of about half the proteolytic sites can occur without a decrease in maximal proteolytic rate (Lee *et al*, 2009). Thus, in Mtb ClpP1P2 partial occupancy of active sites by activators probably could occur without reducing activity, while also perhaps inducing conformational changes in remaining subunits that favour formation of the active state. To induce these structural changes, the activators exhibit very strong cooperativity with a Hill coefficient between 5 and 7, which suggests that multiple molecules bind to either a fraction of the active sites (or to a distinct allosteric site) to induce the active conformation. Unfortunately, the activators identified thus far have low affinities; consequently, identifying their specific binding sites is difficult by standard biochemical approaches.

Although these activators resemble peptide substrates, there does not appear to be a simple correspondence between sequences that are preferentially hydrolysed and ones that support activation. Peptides corresponding to the peptide activators were poor substrates for the enzyme (Table I), and conversely the peptide corresponding to the best substrate was poor as an activator. Thus, it is possible that these peptides activate by also binding to an additional regulatory site.

The active sites of the cylindrical proteases, such as ClpPs, HsUV, or proteasomes, are sequestered within the proteolytic chamber and by themselves cannot degrade protein substrates (Baumeister *et al*, 1998; Yu and Houry, 2007; Striebel *et al*, 2009). Activation of these compartmentalized proteases can occur if the binding of the ATPase alters the conformation of the active site as in HslUV system (Yoo *et al*, 1996; Huang and Goldberg, 1997; Sousa *et al*, 2002; Yu and Houry, 2007) or opens an entry channel to allow substrate access, as occurs in gating of proteasome (Groll *et al*, 2000; Whitby *et al*, 2000; Smith *et al*, 2007; Rabl *et al*, 2008) and proteases ClpXP or ClpAP (Grimaud *et al*, 1998; Maurizi *et al*, 1998; Kirstein *et al*, 2009; Lee *et al*, 2010b). The latter mechanism is important in action of acyldepsipeptide antibiotics, which are cytotoxic in *B. subtilis* and *E. coli* (Brotz-Oesterhelt *et al*, 2005; Kirstein *et al*, 2009) by causing activation of ClpP and excessive degradation of cellular proteins (Kirstein *et al*, 2009; Lee *et al*, 2010a). These molecules bind to the two ends of the ClpP tetradecamer in the cavities between adjacent ClpP monomers, which are the same sites to which loops of the regulatory ATPases bind (Kim *et al*, 2001; Bewley *et al*, 2006; Lee *et al*, 2009, 2010a). Thus, the acyldepsipeptides stimulate proteolysis by facilitating substrate access to the degradative chamber (Kirstein *et al*, 2009; Lee *et al*, 2010a), and prevent association of the protease with regulatory ATPases (Kirstein *et al*, 2009; Lee *et al*, 2010a). Recently, additional activators have been identified that function by a similar mechanism (Leung *et al*, 2011).

One initially attractive possible mechanism of our dipeptide activators was that they function in a similar fashion as acyldepsipeptide antibiotics. However, the dipeptide activator and ClpC1 ATPase have very different effects on peptide degradation, and stimulation of casein degradation requires the presence of both an activator and ClpC1. Furthermore, in related studies using acyldepsipeptide ADEP2, we recently found that these agents also can activate Mtb ClpP1P2 only in the presence of an activating peptide. It is therefore highly unlikely that the activator binds to the same regulatory cavities as ClpC1 or the ADEPs.

The physiological relevance of this activation mechanism for ClpP1P2 is an important issue for future study. The finding that ClpC1 functions only in the presence of a small molecule activator suggests that factors with similar activity exist in Mtb and allow ClpP1P2 association and function with ClpC1. In this respect, these activating dipeptides resemble a protein or 'chemical' chaperone that prevents formation of inactive conformations and favours formation of the active enzyme. However, unlike a chaperone, the dipeptide activators have to be continuously present to maintain the active ClpP1P2 complex. These findings imply that Mtb contains endogenous activators, either small molecules or perhaps protein(s), that promote the formation and the maintenance of ClpP1P2 mixed tetradecamers *in vivo*, and the identification of these activators and the isolation of the active complexes from Mtb will be important goals for future work.

In human mitochondrial ClpP (unlike *E. coli* ClpP), the ClpX ATPase complex is necessary not only for substrate recognition and translocation, but also for the formation of the ClpP tetradecamers from inactive heptamers (Kang *et al*, 2005). This action resembles the second stage in the activation mechanism demonstrated here (Figure 7). However, ClpC1 was unable by itself to induce activation of the Mtb enzyme. In Mtb, the expression of ClpC1 is regulated coordinately with ClpP1 and ClpP2 by ClgR factor (Sherrid *et al*, 2010); therefore, it is very likely that ClpP1P2 functions *in vivo* together with ClpC1. Consequently, our finding that ClpC1 ATPase promotes the degradation of casein only in the presence of an activator argues strongly that the activator serves a unique, essential function, and that in Mtb some endogenous factor serves a similar role in promoting assembly of the mixed complex.

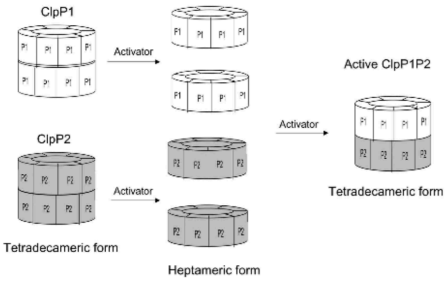


Figure 7 Mechanism of formation of active ClpP1P2. The proposed mechanism is based on the results in (1) Figure 4A showing that the optimal activity of ClpP1P2 is reached at the equal molar amounts of ClpP1 and ClpP2; (2) Figure 3A showing the dissociation of ClpP1 and ClpP2 tetradecamers into heptamers, and their re-association into mixed ClpP1P2 complex in the presence of the activator; (3) Figure 3B showing physical interaction between the rings; and (4) Figure 4B demonstrating crosslinking of only one type of subunits within the rings.

ClpP1P2 is an attractive drug target

The present findings and our related *in-vivo* observations (Raju *et al*, 2011) provide strong evidence that inhibition of ClpP1P2 is a promising new approach to combat tuberculosis. Not only is ClpP1P2 essential in Mtb, but no similar proteolytic complex exists in the mammalian cytosol, and its cleavage specificity, as defined with model peptides, clearly differs from those of the major mammalian cytosolic proteases (proteasomes). Furthermore, Mtb ClpP1P2 in structure and substrate specificity differs considerably from known ClpP family members in the mitochondria. Also, despite its unusual activation mechanism, ClpP1P2 is sensitive to typical inhibitors of serine proteases. It also seems likely that agents that activate Mtb ClpP1P2 may have therapeutic applications, since acyldepsipeptide antibiotics are toxic in many bacteria by activating ClpP and causing excessive proteolysis (Kirstein *et al*, 2009). Finally, the identification of inhibitors of ClpP1P2 activity or blockers of its activation should now be a straightforward undertaking, since screening can be carried out without the need for ATPases that regulate its activity *in vivo*. Nevertheless, characterization of these essential regulatory ATPases will be an important step to understand the physiological functions of ClpP1P2, and they may also represent therapeutic targets. In fact, Mtb ClpC1 has been recently identified as a drug target (Schmitt *et al*, 2011).

Materials and methods

Materials

Fluorogenic peptide substrates with C-terminal aminomethylcoumarin (amc) groups and protease inhibitor were from Bachem (Switzerland), Sigma (USA), or Biomol International (USA). Peptide substrates for the FRET assay, FITC-casein, and glutaraldehyde were from Sigma. Ni-NTA agarose was from Qiagen and Sephacryl S-300 from Pharmacia. Black 96-well microplates used in the enzyme assays were from Greiner (Germany). FP-Biotin, 10-(fluoroethoxyphosphinyl)-N-(biotinamidopentyl)decanamide, was a gift of Dr Francesco Parlati (Onyx Inc). ADEP2 was kindly provided by Dr Heike Brötz-Oesterhelt (Heinrich-Heine-University Düsseldorf).

Bacterial strains, plasmids, expression and growth of cells

M. smegmatis mc^{2}155 were grown at 37°C in Middlebrook 7H9 broth with 0.05% Tween-80 and ADC (0.5% BSA, 0.2% dextrose, 0.085% NaCl, 0.003 g catalase/1 L media) supplemented with hygromycin (50 μg/ml) and in case of inducible expression of ClpP1, ClpP2, or ClpC1, with anhydrotetracycline (100 ng/ml). Full-length C-terminally 6 × His-, Myc- or 6 × His and Myc-tagged ClpP1 and ClpP2 proteins were expressed in *M. smegmatis* on pMV plasmid under the regulation of a constitutive GroEL promoter. For expression of shorter forms of polypeptides corresponding to processed ClpP1 (lacking 6 N-terminal amino acids) and ClpP2 (lacking 11 N-terminal amino acids), pTetOR plasmid, which has an inducible tetracycline promoter was used. After overnight induction with ATc (100 ng/ml), cells were collected and kept at −70°C.

Purification of Mtb ClpP1 and ClpP2

All procedures for enzyme purification were carried out at 4°C using the following buffers:

Buffer A: 50 mM potassium phosphate pH 7.6 100 mM KCl, 5 mM $MgCl_2$, β-mercaptoethanol, and 10% glycerol; *buffer B*: 50 mM potassium phosphate pH 7.6 100 mM KCl, 5 mM $MgCl_2$, and 5% glycerol. In a typical purification, frozen cells (5–10 g) expressing ClpP1 or ClpP2 were suspended in two volumes of buffer A and broken by French press at 1500 p.s.i. The extract was centrifuged at 100 000 *g* and mixed with 5 ml Ni-NTA agarose previously equilibrated in buffer A. After incubating at 4°C for 4 h, Ni-NTA agarose resin was transferred to empty column and proteins were eluted using step gradient (0, 25, 50, 100, and 200 mM of imidazole in buffer B). The fractions containing near homogeneous ClpP1 or ClpP2 proteins were combined, concentrated on Millipore MWCO 10 000 cut filter and purified further by gel filtration on Sephacryl S-300 column (1.5 × 12 cm) equilibrated with buffer B. The specific activity of the purified enzyme was in the range 4–5 μmole/mg/min High molecular weight protein peaks were combined, concentrated to ∼2–5 mg/ml, and kept at −80°C.

Determination of enzymatic activities

All assays of a proteolytic activity were performed at 37°C in 96-well plate using Plate Reader SpectraMax M5 (Molecular Devices, USA). Wells contained 0.1 mM fluorescent peptide, 0.3–3 μg ClpP1P2, 0.5 mM Z-Leu-Leu-aldehyde or 5 mM Z-Leu-Leu in 80 μl of 50 mM phosphate buffer pH 7.6 with 5% glycerol and 100 mM KCl. After shaking for 20 s, peptidase activities were assayed at 37°C by continuously monitoring the rate of production of fluorescent 7-amino-4-methylcoumarin (AMC) from fluorogenic peptide substrates at 460 nm (Ex at 380 nm). Cleavage of longer octa- and nano-peptides was measured using FRET assay using the same conditions, except that 0.1–0.5 μg enzyme and 2–5 mM quenched substrates were used. Increase in fluorescence was monitored continuously at 405 nm (Ex at 340 nm). FITC-casein was purified using PD-10 column, and its hydrolysis (2–5 μg) was continuously monitored at 518 nm (Ex at 492 nm). All assays were performed in triplicate and average results presented. Deviations in the measurements for AMC and quench peptide substrates were <5%, while in the case of FITC-casein, it was <10%. Potential cleavage of enzyme activator Z-Leu-Leu was tested by fluorescamine method as described previously (Akopian *et al*, 1997).

ATPase activity of ClpC1 was measured in the buffer containing 50 mM Tris–HCl pH 7.8, 50 mM KCl, 10% glycerol, 1 mM DTT, 2 mM ATP, and 8 mM $MgCl_2$. The amount of generated orthophosphate was measured colorimetrically by Malachite Green method (Baykov *et al*, 1988). The deviation between the measurements was <5%.

Fluorescent emission spectra

Emission spectra of ClpP1 complexes were determined in buffer B in microplate format using a SpectroMax M5 (USA) plate reader with a step of 1 nm. Preliminarily spectra registration indicated a max of excitation of Trp174 in ClpP1 environment is 279 nm in buffer B.

MS and N-terminal analysis

Protein bands from SDS–PAGE were digested by sequence grade trypsin (Promega) or chymotrypsin (Roche). Obtained peptides were analysed by Thermo Electron LTQ-Orbitrap MassSpec after their separation by Agilent 1100 HPLC system. Automatic N-terminal sequencing of purified proteins (Edman degradation)

was done using ABI 494 Protein Sequencers after transfer of proteins onto PVDF membranes.

Crosslinking of neighbouring subunits

Crosslinking of ClpP1P2 was carried out with 0.125% glutaraldehyde in the buffer B containing activator Z-Leu-Leu. After 0.5 and 20 h incubation at room temperature, the reaction mixture was resolved by SDS–PAGE, and proteins analysed by MS.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

These studies were supported by NIGMS 5R01 GM51923-16 and Harvard Catalyst Pilot Grants to ALG.

Author contributions: TA, OK, and ALG conceived and designed experiments; TA, OK, RR, and MU performed experiments; TA, OK, ER, and ALG analysed the data; TA, OK, and ALG wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Akopian TN, Kisselev AF, Goldberg AL (1997) Processive degradation of proteins and other catalytic properties of the proteasome from Thermoplasma acidophilum. *J Biol Chem* **272:** 1791–1798

Andersson FI, Tryggvesson A, Sharon M, Diemand AV, Classen M, Best C, Schmidt R, Schelin J, Stanne TM, Bukau B, Robinson CV, Witt S, Mogk A, Clarke AK (2009) Structure and function of a novel type of ATP-dependent Clp protease. *J Biol Chem* **284:** 13519–13532

Barik S, Sureka K, Mukherjee P, Basu J, Kundu M (2010) RseA, the SigE specific anti-sigma factor of Mycobacterium tuberculosis, is inactivated by phosphorylation-dependent ClpC1P2 proteolysis. *Mol Microbiol* **75:** 592–606

Baumeister W, Walz J, Zühl F, Seemüller E (1998) The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92:** 367–380

Baykov AA, Evtushenko OA, Avaeva SM (1988) A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal Biochem* **171:** 266–270

Benaroudj N, Raynal B, Miot M, Ortiz-Lombardia M (2011) Assembly and proteolytic processing of mycobacterial ClpP1 and ClpP2. *BMC Biochem* **12:** 61

Bewley MC, Graziano V, Griffin K, Flanagan JM (2006) The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. *J Struct Biol* **153:** 113–128

Brotz-Oesterhelt H, Beyer D, Kroll HP, Endermann R, Ladel C, Schroeder W, Hinzen B, Raddatz S, Paulsen H, Henninger K, Bandow JE, Sahl HG, Labischinski H (2005) Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nat Med* **11:** 1082–1087

Butler SM, Festa RA, Pearce MJ, Darwin KH (2006) Self-compartmentalized bacterial proteases and pathogenesis. *Mol Microbiol* **60:** 553–562

Flanagan JM, Wall JS, Capel MS, Schneider DK, Shanklin J (1995) Scanning transmission electron microscopy and small-angle scattering provide evidence that native Escherichia coli ClpP is a tetradecamer with an axial pore. *Biochemistry* **34:** 10910–10917

Grimaud R, Kessel M, Beuron F, Steven AC, Maurizi MR (1998) Enzymatic and structural similarities between the Escherichia coli ATP-dependent proteases, ClpXP and ClpAP. *J Biol Chem* **273:** 12476–12481

Groll M, Bajorek M, Kohler A, Moroder L, Rubin DM, Huber R, Glickman MH, Finley D (2000) A gated channel into the proteasome core particle. *Nat Struct Biol* **7:** 1062–1067

Hoskins JR, Pak M, Maurizi MR, Wickner S (1998) The role of the ClpA chaperone in proteolysis by ClpAP. *Proc Natl Acad Sci USA* **95:** 12135–12140

Huang H, Goldberg AL (1997) Proteolytic activity of the ATP-dependent protease HslVU can be uncoupled from ATP hydrolysis. *J Biol Chem* **272:** 21364–21372

Hwang BJ, Park WJ, Chung CH, Goldberg AL (1987) Escherichia coli contains a soluble ATP-dependent protease (Ti) distinct from protease La. *Proc Natl Acad Sci USA* **84:** 5550–5554

Hwang BJ, Woo KM, Goldberg AL, Chung CH (1988) Protease Ti, a new ATP-dependent protease in Escherichia coli, contains protein-activated ATPase and proteolytic functions in distinct subunits. *J Biol Chem* **263:** 8727–8734

Ingvarsson H, Mate MJ, Hogbom M, Portnoi D, Benaroudj N, Alzari PM, Ortiz-Lombardia M, Unge T (2007) Insights into the inter-ring plasticity of caseinolytic proteases from the X-ray structure of Mycobacterium tuberculosis ClpP1. *Acta Crystallogr D Biol Crystallogr* **63:** 249–259

Ishikawa T, Beuron F, Kessel M, Wickner S, Maurizi MR, Steven AC (2001) Translocation pathway of protein substrates in ClpAP protease. *Proc Natl Acad Sci USA* **98:** 4328–4333

Kang SG, Dimitrova MN, Ortega J, Ginsburg A, Maurizi MR (2005) Human mitochondrial ClpP is a stable heptamer that assembles into a tetradecamer in the presence of ClpX. *J Biol Chem* **280:** 35424–35432

Katayama-Fujimura Y, Gottesman S, Maurizi MR (1987) A multiple-component, ATP-dependent protease from Escherichia coli. *J Biol Chem* **262:** 4477–4485

Kim YI, Levchenko I, Fraczkowska K, Woodruff RV, Sauer RT, Baker TA (2001) Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat Struct Biol* **8:** 230–233

Kirstein J, Hoffmann A, Lilie H, Schmidt R, Rubsamen-Waigmann H, Brotz-Oesterhelt H, Mogk A, Turgay K (2009) The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. *EMBO Mol Med* **1:** 37–49

Kisselev AF, Callard A, Goldberg AL (2006) Importance of the different proteolytic sites of the proteasome and the efficacy of inhibitors varies with the protein substrate. *J Biol Chem* **281:** 8582–8590

Kisselev AF, Garcia-Calvo M, Overkleeft HS, Peterson E, Pennington MW, Ploegh HL, Thornberry NA, Goldberg AL (2003) The QJ;caspase-like sites of proteasomes, their substrate specificity, new inhibitors and substrates, and allosteric interactions with the trypsin-like sites. *J Biol Chem* **278:** 35869–35877

Kress W, Maglica Z, Weber-Ban E (2009) Clp chaperone-proteases: structure and function. *Res Microbiol* **160:** 618–628

Lee BG, Park EY, Lee KE, Jeon H, Sung KH, Paulsen H, Rubsamen-Schaeff H, Brotz-Oesterhelt H, Song HK (2010a) Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nat Struct Mol Biol* **17:** 471–478

Lee JW, Park E, Jeong MS, Jeon YJ, Eom SH, Seol JH, Chung CH (2009) HslVU ATP-dependent protease utilizes maximally six among twelve threonine active sites during proteolysis. *J Biol Chem* **284:** 33475–33484

Lee ME, Baker TA, Sauer RT (2010b) Control of substrate gating and translocation into ClpP by channel residues and ClpX binding. *J Mol Biol* **399:** 707–718

Leung E, Datti A, Cossette M, Goodreid J, McCaw SE, Mah M, Nakhamchik A, Ogata K, El Bakkouri M, Cheng YQ, Wodak SJ, Eger BT, Pai EF, Liu J, Gray-Owen S, Batey RA, Houry WA (2011) Activators of cylindrical proteases as antimicrobials: identification and development of small molecule activators of ClpP protease. *Chem Biol* **18:** 1167–1178

Liu Y, Patricelli MP, Cravatt BF (1999) Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci USA* **96:** 14694–14699

Maurizi MR (1991) ATP-promoted interaction between Clp A and Clp P in activation of Clp protease from Escherichia coli. *Biochem Soc Trans* **19:** 719–723

Maurizi MR, Clark WP, Katayama Y, Rudikoff S, Pumphrey J, Bowers B, Gottesman S (1990a) Sequence and structure of Clp

P, the proteolytic component of the ATP-dependent Clp protease of Escherichia coli. *J Biol Chem* **265:** 12536–12545

Maurizi MR, Clark WP, Kim SH, Gottesman S (1990b) Clp P represents a unique family of serine proteases. *J Biol Chem* **265:** 12546–12552

Maurizi MR, Singh SK, Thompson MW, Kessel M, Ginsburg A (1998) Molecular properties of ClpAP protease of Escherichia coli: ATP-dependent association of ClpA and clpP. *Biochemistry* **37:** 7778–7786

Maurizi MR, Thompson MW, Singh SK, Kim SH (1994) Endopeptidase Clp: ATP-dependent Clp protease from Escherichia coli. *Methods Enzymol* **244:** 314–331

Ollinger J, O'Malley T, Kesicki EA, Odingo J, Parish T (2012) Validation of the essential ClpP protease in Mycobacterium tuberculosis as a novel drug target. *J Bacteriol* **194:** 663–668

Ortega J, Singh SK, Ishikawa T, Maurizi MR, Steven AC (2000) Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Mol Cell* **6:** 1515–1521

Peltier JB, Ripoll DR, Friso G, Rudella A, Cai Y, Ytterberg J, Giacomelli L, Pillardy J, van Wijk KJ (2004) Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. *J Biol Chem* **279:** 4768–4781

Peltier JB, Ytterberg J, Liberles DA, Roepstorff P, van Wijk KJ (2001) Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of Arabidopsis thaliana. *J Biol Chem* **276:** 16318–16327

Porankiewicz J, Wang J, Clarke AK (1999) New insights into the ATP-dependent Clp protease: Escherichia coli and beyond. *Mol Microbiol* **32:** 449–458

Powers JC, Kam CM (1994) Isocoumarin inhibitors of serine peptidases. *Methods Enzymol* **244:** 442–457

Rabl J, Smith DM, Yu Y, Chang SC, Goldberg AL, Cheng Y (2008) Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol Cell* **30:** 360–368

Raju R, Unnikrishnan M, Rubin D, Krishnamoorthy V, Kandror O, Akopian T, Goldberg AL, Rubin EJ (2011) Mycobacterium tuberculosis encodes a heteromeric ClpP protease required for normal growth and virulence. *PLoS Pathog* (in press)

Reid BG, Fenton WA, Horwich AL, Weber-Ban EU (2001) ClpA mediates directional translocation of substrate proteins into the ClpP protease. *Proc Natl Acad Sci USA* **98:** 3768–3772

Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48:** 77–84

Schelin J, Lindmark F, Clarke AK (2002) The clpP multigene family for the ATP-dependent Clp protease in the cyanobacterium Synechococcus. *Microbiology* **148:** 2255–2265

Schmitt EK, Riwanto M, Sambandamurthy V, Roggo S, Miault C, Zwingelstein C, Krastel P, Noble C, Beer D, Rao SP, Au M, Niyomrattanakit P, Lim V, Zheng J, Jeffery D, Pethe K, Camacho LR (2011) The natural product cyclomarin kills Mycobacterium tuberculosis by targeting the ClpC1 subunit of the caseinolytic protease. *Angew Chem Int Ed Engl* **50:** 5889–5891

Seol JH, Yoo SJ, Shin DH, Shim YK, Kang MS, Goldberg AL, Chung CH (1997) The heat-shock protein HslVU from Escherichia coli is a protein-activated ATPase as well as an ATP-dependent proteinase. *Eur J Biochem* **247:** 1143–1150

Sherrid AM, Rustad TR, Cangelosi GA, Sherman DR (2010) Characterization of a Clp protease gene regulator and the reaeration response in Mycobacterium tuberculosis. *PLoS One* **5:** e11622

Shin DH, Lee CS, Chung CH, Suh SW (1996) Molecular symmetry of the ClpP component of the ATP-dependent Clp protease, an Escherichia coli homolog of 20 S proteasome. *J Mol Biol* **262:** 71–76

Sjogren LL, Stanne TM, Zheng B, Sutinen S, Clarke AK (2006) Structural and functional insights into the chloroplast ATP-dependent Clp protease in Arabidopsis. *Plant Cell* **18:** 2635–2649

Smith DM, Chang SC, Park S, Finley D, Cheng Y, Goldberg AL (2007) Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry. *Mol Cell* **27:** 731–744

Smith N, Naude RJ, Oelofsen W, Lazure C, Patthy A (1992) The isolation and partial characterization of trypsinogen, pancreatic secretory trypsin inhibitor and multiple forms of chymotrypsinogen and trypsin from the pancreas of the ostrich (Struthio camelus). *Int J Biochem* **24:** 877–885

Sousa MC, Kessler BM, Overkleeft HS, McKay DB (2002) Crystal structure of HslUV complexed with a vinyl sulfone inhibitor: corroboration of a proposed mechanism of allosteric activation of HslV by HslU. *J Mol Biol* **318:** 779–785

Stanne TM, Pojidaeva E, Andersson FI, Clarke AK (2007) Distinctive types of ATP-dependent Clp proteases in cyanobacteria. *J Biol Chem* **282:** 14394–14402

Striebel F, Kress W, Weber-Ban E (2009) Controlled destruction: AAA+ ATPases in protein degradation from bacteria to eukaryotes. *Curr Opin Struct Biol* **19:** 209–217

Szyk A, Maurizi MR (2006) Crystal structure at 1.9A of E. coli ClpP with a peptide covalently bound at the active site. *J Struct Biol* **156:** 165–174

Thompson MW, Maurizi MR (1994) Activity and specificity of Escherichia coli ClpAP protease in cleaving model peptide substrates. *J Biol Chem* **269:** 18201–18208

Viala J, Mazodier P (2002) ClpP-dependent degradation of PopR allows tightly regulated expression of the clpP3 clpP4 operon in Streptomyces lividans. *Mol Microbiol* **44:** 633–643

Viala J, Rapoport G, Mazodier P (2000) The clpP multigenic family in Streptomyces lividans: conditional expression of the clpP3 clpP4 operon is controlled by PopR, a novel transcriptional activator. *Mol Microbiol* **38:** 602–612

Wang J, Hartling JA, Flanagan JM (1997) The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell* **91:** 447–456

Waxman L, Goldberg AL (1982) Protease La from Escherichia coli hydrolyzes ATP and proteins in a linked fashion. *Proc Natl Acad Sci USA* **79:** 4883–4887

Whitby FG, Masters EI, Kramer L, Knowlton JR, Yao Y, Wang CC, Hill CP (2000) Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* **408:** 115–120

Yoo SJ, Seol JH, Shin DH, Rohrwild M, Kang MS, Tanaka K, Goldberg AL, Chung CH (1996) Purification and characterization of the heat shock proteins HslV and HslU that form a new ATP-dependent protease in Escherichia coli. *J Biol Chem* **271:** 14035–14040

Yu AY, Houry WA (2007) ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Lett* **581:** 3749–3757

Appendix 2 – Over and under represented proteins identified through proteomic profiling of P750-clpP1P2DAS Mtb in the presence (mutant) and absence (wildtype) of ATc

The tables below denote significantly over and under represented proteins identified from our quantitative proteomics screen to identify potential protease substrates that accumulated upon the depletion of Clp in Mtb. For each protein, the following are listed: Rv number, protein name, number of peptides identified, normalized summed intensity for each TMT channel (wildtype: TMT126-TMT128; mutant: TMT129-TMT131), the average ratio of mutant/wildtype (mut/wt) protein intensity, and the p-value as determined by a t-test grouping the three biological replicates for each condition. Proteins are sorted from the largest to smallest average mut/wt ratio.

OVER-REPRESENTED PROTEINS

RV	Protein	# pep.	126 (wt)	127 (wt)	128 (wt)	129 (mut)	130 (mut)	131 (mut)	mut/wt ratio	p-value
Rv0251c	heat shock protein Hsp20	22	0.61	0.43	0.26	31.45	35.32	31.93	75.92	0.0013
Rv1741	hypothetical protein Rv1741	1	1.96	1.51	0.61	32.59	30.78	32.55	23.51	0.0000
Rv2979c	resolvase	1	2.21	1.39	1.85	31.31	30.59	32.64	17.35	0.0001
Rv2031c	heat shock protein hspX	28	1.90	1.94	1.66	31.24	31.00	32.27	17.18	0.0001
Rv1915	isocitrate lyase	4	2.25	2.26	1.79	32.47	29.29	31.93	14.87	0.0009
Rv0605	resolvase	6	2.09	2.12	2.20	29.55	31.78	32.26	14.60	0.0008
Rv1072	transmembrane protein	14	2.38	2.87	1.64	35.02	25.07	33.01	13.51	0.0102
Rv2710	RNA polymerase sigma factor SigB	12	3.17	3.05	2.22	33.75	27.88	29.93	10.85	0.0031
Rv3334	MerR family transcriptional regulator	2	2.62	3.82	2.15	31.86	22.77	36.79	10.64	0.0202
Rv1916	isocitrate lyase	2	3.17	2.87	2.77	25.38	30.89	34.92	10.35	0.0099
Rv2079	hypothetical protein Rv2079	2	3.13	2.70	2.98	31.55	28.94	30.70	10.35	0.0006
Rv2791c	transposase	7	2.99	3.33	2.57	29.45	28.36	33.30	10.25	0.0025
Rv1765c	hypothetical protein Rv1765c	1	3.74	3.34	2.61	29.14	31.81	29.36	9.32	0.0002
Rv2107	PE family protein	2	3.58	3.63	2.79	26.70	29.85	33.46	9.00	0.0047
Rv2917	hypothetical protein Rv2917	1	3.36	3.45	3.36	29.29	29.98	30.56	8.83	0.0002
Rv3408	hypothetical protein Rv3408	1	3.19	4.47	2.51	30.09	22.86	36.87	8.83	0.0207
Rv0192	hypothetical protein Rv0192	2	3.96	3.33	3.00	26.02	30.21	33.48	8.72	0.0059
Rv2652c	phiRv2 prophage protein	1	4.20	3.65	3.88	28.33	28.07	31.86	7.52	0.0020
Rv2850c	magnesium chelatase	1	4.69	3.34	4.42	28.09	29.29	30.17	7.03	0.0000
Rv1945	hypothetical protein Rv1945	1	4.46	4.61	3.40	30.48	28.60	28.45	7.02	0.0000
Rv3053c	glutaredoxin electron transport protein NrdH	4	5.42	4.30	2.91	32.20	25.19	29.99	6.92	0.0033
Rv1128c	hypothetical protein Rv1128c	3	3.86	4.44	4.41	25.92	31.05	30.32	6.87	0.0037
Rv0369c	membrane oxidoreductase	1	4.74	3.77	4.42	28.57	30.58	27.92	6.73	0.0003
Rv2016	hypothetical protein Rv2016	4	5.75	4.89	4.59	23.85	30.86	30.05	5.57	0.0078
Rv0220	esterase LipC	2	5.65	4.93	4.70	27.77	28.28	28.67	5.54	0.0000
Rv3769	hypothetical protein Rv3769	2	8.93	4.21	3.17	24.31	28.88	30.49	5.13	0.0009
Rv0945	short chain dehydrogenase	5	5.75	6.45	4.64	30.86	22.89	29.41	4.94	0.0095
Rv2688c	antibiotic ABC transporter ATP-binding protein	3	6.85	5.59	5.55	25.52	28.46	28.02	4.56	0.0003
Rv1471	thioredoxin TRXB1	8	5.64	7.95	4.66	29.34	21.37	31.05	4.48	0.0126
Rv0425c	metal cation transporting P-type ATPase CtpH	1	6.49	6.07	5.73	26.93	24.74	30.04	4.47	0.0046
Rv1854c	NADH dehydrogenase	16	5.68	6.20	7.44	23.96	27.09	29.63	4.18	0.0034
Rv1169c	PE family protein	1	7.04	6.96	5.36	23.81	28.14	28.69	4.17	0.0026
Rv0984	pterin-4-alpha-carbinolamine dehydratase MoaB2	1	5.15	10.41	3.89	20.16	28.41	31.98	4.14	0.0130
Rv1148c	hypothetical protein Rv1148c	6	7.14	6.41	6.10	26.29	26.63	27.43	4.09	0.0000
Rv2601A	hypothetical protein Rv2601A	1	7.74	7.08	5.19	23.02	28.72	28.26	4.00	0.0033
Rv0275c	TetR family transcriptional regulator	2	6.29	7.12	6.61	28.25	24.80	26.93	4.00	0.0016
Rv0516c	hypothetical protein Rv0516c	3	6.29	6.92	6.82	24.88	28.52	26.57	3.99	0.0021
Rv2324	AsnC family transcriptional regulator	2	7.43	6.36	6.51	24.57	27.02	28.11	3.93	0.0013
Rv0384c	endopeptidase ATP binding protein	112	7.10	6.74	6.58	25.32	27.47	26.78	3.90	0.0006
Rv2431c	PE family protein	2	6.33	7.32	6.87	25.39	28.21	25.88	3.87	0.0008
Rv1431	hypothetical protein Rv1431	2	6.83	6.96	6.99	24.27	26.69	28.25	3.81	0.0035
Rv1343c	lipoprotein LprD	1	9.55	5.47	6.19	27.22	25.33	26.24	3.71	0.0013
Rv1256c	cytochrome P450 130 CYP130	1	7.67	7.06	6.59	24.43	27.26	26.98	3.69	0.0008
Rv0336	13E12 repeat family protein	2	7.04	7.34	6.99	24.36	25.31	28.96	3.68	0.0051
Rv0663	arylsulfatase AtsD	5	6.74	7.40	7.51	25.00	25.37	27.96	3.62	0.0014
Rv2218	lipoyl synthase	2	7.41	7.71	6.54	26.33	24.70	27.31	3.62	0.0003
Rv1182	polyketide synthase associated protein PapA3	1	7.36	8.05	6.54	25.41	25.33	27.31	3.56	0.0001
Rv1331	ATP-dependent Clp protease adaptor protein ClpS	14	7.97	7.65	6.74	25.38	27.37	24.89	3.47	0.0003
Rv2629	hypothetical protein Rv2629	2	6.92	7.15	8.51	23.52	24.62	29.28	3.43	0.0060
N/A	TetR-off repressor protein	9	7.58	7.70	7.33	24.96	26.56	25.86	3.42	0.0004
Rv0805	hypothetical protein Rv0805	2	7.84	7.58	7.23	24.89	25.28	27.19	3.42	0.0009
Rv3054c	hypothetical protein Rv3054c	5	7.71	8.14	7.25	25.28	23.88	27.75	3.33	0.0027

RV	Protein	# pep.	126 (wt)	127 (wt)	128 (wt)	129 (mut)	130 (mut)	131 (mut)	mut/wt ratio	p-value
Rv3406	dioxygenase	2	8.04	8.16	6.99	25.91	25.85	25.06	3.31	0.0000
Rv3190c	hypothetical protein Rv3190c	7	8.04	7.74	7.73	23.79	25.27	27.44	3.25	0.0033
Rv1814	membrane-bound C-5 sterol desaturase erg3 (sterol-c5-desaturase)	1	8.44	7.51	7.74	25.38	25.60	25.34	3.22	0.0001
Rv3884c	CBXX/CFQX family protein	1	8.58	6.96	8.43	24.73	25.42	25.88	3.17	0.0000
Rv0197	oxidoreductase	2	8.66	6.51	8.81	22.88	25.80	27.33	3.17	0.0011
Rv2496c	pyruvate dehydrogenase E1 component beta subunit PdhB	5	7.68	8.66	8.05	25.70	24.80	25.12	3.10	0.0000
Rv2182c	1-acylglycerol-3-phosphate O-acyltransferase	3	7.29	9.30	7.88	29.11	19.18	27.26	3.09	0.0269
Rv1123c	peroxidase BpoB	3	10.01	7.44	7.19	27.07	24.20	24.09	3.06	0.0002
Rv0892	monooxygenase	5	8.15	8.09	8.61	24.50	22.15	28.50	3.02	0.0115
Rv3052c	ribonucleotide reductase stimulatory protein	2	8.38	8.78	7.79	23.16	26.31	25.57	3.01	0.0016
Rv2777c	hypothetical protein Rv2777c	3	7.88	7.43	9.67	22.05	26.60	26.36	3.00	0.0026
Rv2900c	formate dehydrogenase H	2	9.60	7.03	8.37	25.80	22.43	26.77	3.00	0.0013
Rv1047	transposase	1	8.22	7.79	9.15	24.78	24.75	25.32	2.97	0.0001
Rv0938	ATP-dependent DNA ligase	7	8.36	8.36	8.69	22.55	25.60	26.43	2.94	0.0048
Rv3601c	aspartate alpha-decarboxylase	1	12.07	6.39	6.97	27.36	24.79	22.43	2.93	0.0025
Rv0307c	hypothetical protein Rv0307c	2	8.39	7.98	9.09	21.22	24.99	28.33	2.93	0.0137
Rv0507	transmembrane transport protein MmpL2	2	7.91	8.39	9.27	23.03	23.50	27.89	2.91	0.0062
Rv0298	hypothetical protein Rv0298	2	8.22	9.24	8.15	24.55	24.42	25.42	2.90	0.0000
Rv2733c	hypothetical protein Rv2733c	5	8.76	8.47	8.77	23.89	23.82	26.28	2.85	0.0023
Rv3814c	acyltransferase	3	8.65	8.40	9.02	22.29	26.30	25.34	2.84	0.0049
Rv1322A	hypothetical protein Rv1322A	1	8.98	8.85	8.34	24.36	24.19	25.28	2.82	0.0000
Rv0726c	hypothetical protein Rv0726c	1	11.85	7.06	7.79	24.55	26.80	21.96	2.75	0.0016
Rv0103c	cation-transporter P-type ATPase B	5	7.68	9.42	9.88	21.16	24.06	27.81	2.71	0.0091
Rv2633c	hypothetical protein Rv2633c	2	8.91	9.43	8.69	24.18	25.42	23.37	2.70	0.0005
Rv3051c	ribonucleotide-diphosphate reductase subunit alpha	53	9.56	8.04	9.52	23.17	25.34	24.37	2.69	0.0001
Rv0959	hypothetical protein Rv0959	2	9.38	8.68	9.24	23.81	25.11	23.79	2.66	0.0001
Rv0920c	transposase	2	8.60	8.44	10.38	21.94	24.61	26.03	2.65	0.0015
Rv0058	replicative DNA helicase	5	9.14	9.36	8.92	23.24	24.54	24.79	2.65	0.0005
Rv0430	hypothetical protein Rv0430	1	8.54	10.25	8.94	20.55	22.84	28.89	2.61	0.0231
Rv2816c	hypothetical protein Rv2816c	1	8.28	10.61	9.14	22.27	23.95	25.74	2.57	0.0005
Rv0569	hypothetical protein Rv0569	2	9.62	10.32	8.15	25.01	22.38	24.51	2.56	0.0002
Rv2251	flavoprotein	2	9.60	8.87	9.65	21.42	25.28	25.18	2.56	0.0059
Rv3852	histone-like protein HNS	2	7.63	11.74	8.84	22.26	22.42	27.10	2.54	0.0024
Rv0004	hypothetical protein Rv0004	2	10.14	8.91	9.25	21.94	25.74	24.02	2.53	0.0029
Rv2830c	hypothetical protein Rv2830c	3	10.18	8.69	9.47	22.18	25.93	23.54	2.53	0.0023
Rv0386	LuxR/UHPA family transcriptional regulator	1	8.77	11.25	8.40	30.03	19.63	21.92	2.52	0.0366
Rv0233	ribonucleotide-diphosphate reductase subunit beta	1	9.31	10.24	8.93	25.38	21.19	24.95	2.51	0.0053
Rv1828	hypothetical protein Rv1828	1	10.31	9.81	8.50	24.40	23.31	23.66	2.49	0.0001
Rv1207	dihydropteroate synthase 2 FolP2	2	9.48	9.75	9.41	21.95	23.63	25.79	2.49	0.0057
Rv0486	mannosyltransferase	1	9.59	9.56	9.68	25.07	22.09	24.01	2.47	0.0037
Rv3241c	hypothetical protein Rv3241c	5	10.25	9.44	9.19	22.48	25.43	23.21	2.46	0.0016
Rv0327c	cytochrome P450 135A1	1	10.09	9.39	9.44	21.69	23.80	25.58	2.46	0.0049
Rv2466c	hypothetical protein Rv2466c	9	10.34	8.95	9.73	23.50	24.47	23.01	2.45	0.0000
Rv0426c	hypothetical protein Rv0426c	7	10.30	9.64	9.19	24.78	23.86	22.23	2.43	0.0007
Rv2250A	flavoprotein	5	10.16	8.92	10.09	22.91	24.04	23.88	2.43	0.0000
Rv2933	phenolpthiocerol synthesis type-I polyketide synthase PPSC	3	9.04	9.12	11.30	19.24	24.81	26.49	2.39	0.0166
Rv1129c	transcriptional regulator protein	2	10.05	8.87	10.54	21.25	25.56	23.73	2.39	0.0036
Rv1956	transcriptional regulatory protein	6	10.28	11.26	8.00	23.60	23.21	23.64	2.38	0.0043
Rv2017	transcriptional regulatory protein	1	10.10	9.91	9.54	27.11	20.55	22.79	2.38	0.0188
Rv3815c	acyltransferase	1	6.69	11.95	11.10	26.22	16.93	27.11	2.36	0.0351

RV	Protein	# pep.	126 (wt)	127 (wt)	128 (wt)	129 (mut)	130 (mut)	131 (mut)	mut/wt ratio	p-value
Rv3130c	triacylglycerol synthase	3	9.56	9.12	11.07	20.74	24.70	24.81	2.36	0.0038
Rv2252	diacylglycerol kinase	5	10.02	9.94	9.89	23.64	22.80	23.71	2.35	0.0004
Rv3378c	hypothetical protein Rv3378c	1	9.84	8.88	11.20	20.92	24.61	24.55	2.34	0.0020
Rv1633	excinuclease ABC subunit B	10	10.06	9.65	10.35	22.11	24.04	23.78	2.33	0.0008
Rv1500	glycosyltransferase	3	9.52	10.11	10.45	21.89	20.46	27.57	2.32	0.0242
Rv0939	bifunctional 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase/cyclase/dehydrase	2	9.88	9.65	10.67	20.87	25.18	23.75	2.31	0.0065
Rv2702	polyphosphate glucokinase PPGK (polyphosphate-glucose phosphotransferase)	2	10.04	9.59	10.78	24.00	23.12	22.46	2.29	0.0000
Rv3768	hypothetical protein Rv3768	3	9.69	10.50	10.25	22.52	22.28	24.76	2.29	0.0019
Rv1265	hypothetical protein Rv1265	4	10.85	9.65	10.01	20.93	25.50	23.05	2.28	0.0069
Rv0781	oligopeptidase B	2	10.03	10.57	9.92	22.89	23.62	22.97	2.28	0.0000
Rv3173c	TetR/ACRR family transcriptional regulator	1	16.25	7.98	6.29	21.32	26.99	21.16	2.28	0.0311
Rv1992c	metal cation transporter P-type ATPase G CtpG	1	9.71	11.64	9.22	26.61	16.34	26.49	2.27	0.0567
Rv3219	transcriptional regulatory protein WHIB-like WHIB1	1	11.35	10.02	9.32	20.22	26.09	23.00	2.26	0.0102
Rv0676c	transmembrane transport protein MmpL5	1	10.77	9.49	10.46	22.35	24.84	22.10	2.26	0.0014
Rv0869c	molybdenum cofactor biosynthesis protein A	2	11.28	9.71	9.74	23.95	22.32	23.00	2.25	0.0001
Rv0274	hypothetical protein Rv0274	7	9.89	10.53	10.36	22.64	22.30	24.27	2.25	0.0010
Rv2823c	hypothetical protein Rv2823c	2	11.13	8.79	10.92	21.40	24.93	22.83	2.24	0.0008
Rv2386c	salicylate synthase MbtI	1	10.33	10.36	10.24	21.97	22.80	24.30	2.23	0.0028
Rv2206	transmembrane protein	6	9.60	9.94	11.40	22.11	22.93	24.02	2.23	0.0001
Rv2050	hypothetical protein Rv2050	1	11.04	10.67	9.26	21.37	22.66	25.00	2.23	0.0018
Rv3116	molybdenum cofactor biosynthesis protein MoeB	5	9.53	9.31	12.24	20.14	24.47	24.31	2.22	0.0030
Rv0350	molecular chaperone DnaK	235	10.63	11.04	9.53	24.52	21.62	22.67	2.21	0.0009
Rv2326c	transmembrane ATP-binding protein ABC transorter	3	10.09	10.71	10.61	21.90	22.24	24.45	2.18	0.0027
Rv0753c	methylmalonate-semialdehyde dehydrogenase	15	9.48	11.61	10.49	22.05	22.20	24.17	2.17	0.0002
Rv0677c	hypothetical protein Rv0677c	3	13.47	7.83	10.65	24.20	21.98	21.87	2.13	0.0081
Rv3874	10 kDa culture filtrate antigen EsxB	2	10.14	11.98	9.88	20.72	21.91	25.37	2.13	0.0053
Rv0885	hypothetical protein Rv0885	2	9.15	10.10	12.81	18.45	23.92	25.58	2.12	0.0162
Rv0752c	acyl-CoA dehydrogenase FADE9	4	11.04	10.07	11.00	21.44	24.32	22.14	2.11	0.0023
Rv1635c	transmembrane protein	2	10.75	9.56	11.81	21.82	23.02	23.06	2.11	0.0003
Rv0776c	hypothetical protein Rv0776c	2	10.26	10.54	11.47	21.05	22.66	24.01	2.10	0.0017
Rv3765c	two component transcriptional regulatory protein	3	10.63	10.05	11.65	21.76	22.95	22.96	2.09	0.0001
Rv1249c	hypothetical protein Rv1249c	3	11.67	10.73	9.93	23.61	21.14	22.92	2.09	0.0004
Rv2559c	recombination factor protein RarA	2	9.70	11.49	11.21	20.96	21.69	24.95	2.09	0.0042
Rv3583c	transcription factor, CarD	18	11.33	11.06	10.02	22.21	23.59	21.79	2.09	0.0001
Rv3526	oxidoreductase	5	14.29	8.17	10.10	26.20	19.81	21.44	2.07	0.0116
Rv3875	6 kDa early secretory antigenic target ESXA (ESAT-6)	10	9.93	10.99	11.66	20.25	19.40	27.76	2.07	0.0443
Rv2108	PPE family protein	3	11.67	10.62	10.32	22.11	23.09	22.19	2.07	0.0000
Rv3627c	hypothetical protein Rv3627c	3	10.42	10.59	11.62	21.99	21.55	23.82	2.06	0.0006
Rv2124c	5-methyltetrahydrofolate--homocystein methyltransferase	35	10.75	10.86	11.23	22.12	21.74	23.30	2.05	0.0007
Rv2276	cytochrome P450 121 CYP121	4	9.84	12.17	10.92	22.20	20.56	24.31	2.04	0.0019
Rv0352	chaperone protein DnaJ1	19	11.10	12.54	9.38	22.66	20.75	23.57	2.03	0.0008
Rv1783	hypothetical protein Rv1783	5	10.57	10.50	12.00	21.97	21.80	23.16	2.02	0.0001
Rv2714	hypothetical protein Rv2714	3	10.44	11.59	11.07	22.32	21.30	23.28	2.02	0.0003
Rv2725c	GTP-binding protein HflX	3	10.81	11.62	10.67	22.45	21.07	23.38	2.02	0.0010
Rv1446c	putative OXPP cycle protein OPCA	6	10.76	10.45	11.89	20.20	23.90	22.79	2.02	0.0042
Rv1784	hypothetical protein Rv1784	7	10.02	11.28	11.91	20.92	22.24	23.63	2.01	0.0005

RV	Protein	# pep.	126 (wt)	127 (wt)	128 (wt)	129 (mut)	130 (mut)	131 (mut)	mut/wt ratio	p-value
Rv1037c	putative ESAT-6 like protein ESXI (ESAT-6 like protein 1)	1	12.68	8.48	12.11	20.72	22.54	23.47	2.01	0.0039
Rv0790c	hypothetical protein Rv0790c	1	9.15	11.76	12.37	22.62	20.14	23.95	2.00	0.0018
Rv2808	hypothetical protein Rv2808	1	12.00	11.36	9.94	24.25	19.19	23.27	2.00	0.0105
Rv1278	hypothetical protein Rv1278	9	9.05	13.48	10.82	21.50	19.74	25.41	2.00	0.0075

UNDER-REPRESENTED PROTEINS

RV	Protein	# pep.	126 (wt)	127 (wt)	128 (wt)	129 (mut)	130 (mut)	131 (mut)	mut/wt ratio	p-value
Rv3618	monooxygenase	1	20.05	21.60	25.16	10.84	11.10	11.25	0.50	0.0172
Rv1130	hypothetical protein Rv1130	2	21.22	21.32	24.36	11.14	9.86	12.10	0.49	0.0016
Rv0987	adhesion component transport transmembrane protein ABC transporter	2	20.96	22.14	24.07	10.82	10.59	11.41	0.49	0.0039
Rv3587c	hypothetical protein Rv3587c	11	22.69	21.57	22.95	10.55	11.71	10.53	0.49	0.0000
Rv0986	adhesion component transport ATP-binding protein ABC transporter	10	20.98	23.81	22.84	11.00	10.27	11.11	0.48	0.0026
Rv0108c	hypothetical protein Rv0108c	1	20.63	28.23	19.29	9.07	10.91	11.87	0.47	0.0400
Rv0467	isocitrate lyase	35	23.49	22.11	22.69	10.55	11.13	10.03	0.46	0.0000
Rv1774	oxidoreductase	2	20.09	22.18	26.21	9.59	10.67	11.26	0.46	0.0154
Rv2590	fatty-acid-CoA ligase	8	21.73	20.90	25.98	9.51	11.12	10.76	0.46	0.0103
Rv1477	invasion protein	5	22.29	23.98	23.06	9.81	10.60	10.27	0.44	0.0002
Rv2243	acyl-carrier-protein S-malonyltransferase	10	22.41	21.96	25.00	9.60	10.60	10.43	0.44	0.0027
Rv0129c	secreted antigen 85-C FBPC (85C) (antigen 85 complex C) (AG58C) (Mycolyl transferase 85C) (fibronectin-binding protein C)	4	21.70	23.93	23.78	9.98	11.50	9.12	0.44	0.0002
Rv2721c	hypothetical protein Rv2721c	31	21.84	23.99	23.85	10.05	10.15	10.12	0.44	0.0027
Rv2240c	hypothetical protein Rv2240c	2	22.18	23.56	23.98	9.54	10.83	9.91	0.43	0.0001
Rv3048c	ribonucleotide-diphosphate reductase subunit beta	9	21.93	21.88	26.04	9.63	10.61	9.91	0.43	0.0084
Rv3825c	polyketide synthase PKS2	2	23.03	20.88	26.02	9.45	10.03	10.59	0.43	0.0097
Rv3873	PPE family protein	2	17.77	25.33	26.87	9.70	9.57	10.76	0.43	0.0397
Rv2429	alkyl hydroperoxide reductase subunit D	4	22.83	22.90	24.28	9.80	10.30	9.89	0.43	0.0005
Rv3310	acid phosphatase	2	21.99	23.55	24.63	8.70	11.07	10.06	0.43	0.0002
Rv1387	PPE family protein	1	22.12	25.91	22.82	7.27	10.05	11.84	0.41	0.0015
Rv1698	hypothetical protein Rv1698	1	23.07	22.11	26.13	9.19	9.29	10.20	0.40	0.0047
Rv3330	penicillin-binding protein DacB1	3	21.46	24.14	26.86	8.82	9.14	9.57	0.38	0.0096
Rv2780	secreted L-alanine dehydrogenase ALD (40 kDa antigen) (TB43)	6	21.16	24.67	26.71	8.22	9.48	9.77	0.38	0.0074
Rv2525c	hypothetical protein Rv2525c	4	29.73	20.59	23.67	9.60	7.99	8.42	0.35	0.0241
Rv1754c	hypothetical protein Rv1754c	1	24.41	27.31	27.98	6.71	7.58	6.02	0.25	0.0009
Rv1566c	inv protein	4	26.35	27.63	26.31	6.73	6.40	6.58	0.25	0.0003
Rv0063	oxidoreductase	3	26.77	25.54	28.17	6.37	6.70	6.45	0.24	0.0012
Rv0888	hypothetical protein Rv0888	1	21.86	31.23	28.26	4.63	6.45	7.56	0.23	0.0114
Rv2350c	membrane-associated phospholipase C	2	24.74	29.15	35.20	3.26	3.22	4.43	0.12	0.0122
Rv2460c	ATP-dependent ClpP2 protease proteolytic subunit	9	26.54	33.69	32.24	2.18	2.59	2.76	0.08	0.0056
Rv2461c	ATP-dependent ClpP1 protease proteolytic subunit	2	27.58	33.25	33.66	1.35	1.92	2.24	0.06	0.0038

Appendix 3 – Over and under represented proteins identified through quantitative proteomics comparing the proteomes of clpP2_ID Msm and wildtype Msm, both inducibly expressing HIV-2 protease

The tables below denote significantly over and under represented proteins identified from our quantitative proteomics screen to identify potential protease substrates that accumulated upon the depletion of ClpP2 in Msm. ClpP2 depletion was achieved with the clpP2_ID Msm strain in which tetracycline (ATc) inducible HIV-2 expression led to degradation of ClpP2 that was modified to bear a C-terminal inducible degradation tag. Wildtype bacteria also containing an ATc nducible copy of the HIV-2 protease were used as a control to which the clpP2_ID proteome was compared.

For each protein, the following are listed: Rv number, protein name, number of peptides identified, normalized median intensity for each TMT channel, and the ratio of mutant/wildtype (mut/wt) protein intensity at 5h post ATc addition. Proteins are sorted from the largest to smallest mut/wt ratio. TMT reagents were used for each condition as follows: TMT126 (wildtype 0h ATc), TMT127 (wildtype 5h ATc), TMT128 (wildtype 11h ATc), TMT129 (clpP2_ID 0h ATc), TMT130 (clpP2_ID 5h ATc), and TMT131 (clpP2_ID 11h ATc).

OVER-REPRESENTED PROTEINS

NCBI ID	Protein	# pep.	TMT126 (wt 0h)	TMT127 (wt 5h)	TMT128 (wt 11h)	TMT129 (clpP2 0h)	TMT130 (clpP2 5h)	TMT131 (clpP2 11h)	clpP2/wt 5h
YP_884588.1	transmembrane protein, putative	2	228.50	304.66	347.69	529.32	5401.97	5320.27	17.73
YP_887885.1	ATP-dependent protease La	9	1500.00	537.64	644.88	1308.32	9124.08	20441.50	16.97
YP_888562.1	lipoyltransferase	2	572.00	104.91	145.66	351.50	1569.43	1988.94	14.96
YP_887343.1	orro	2	3790.00	651.29	568.21	1372.81	5934.92	5119.50	9.11
YP_889521.1	permease of the major facilitator superfamily protein	10	783.50	208.94	816.69	755.51	1888.77	1738.44	9.04
YP_884837.1	Hsp20/alpha crystallin family protein	5	193.00	193.20	244.42	175.98	1603.11	3860.16	8.30
YP_886115.1	Fatty acid desaturase	4	1219.00	476.89	335.07	558.80	3632.58	2244.91	7.62
YP_886285.1	transcription factor WhiB	7	9080.00	1975.73	1434.06	6016.42	13302.40	7619.93	6.73
YP_885824.1	50S ribosomal protein L22	10	176000.00	24871.49	8284.19	77762.02	160310.98	127759.37	6.45
YP_886199.1	F420-0--gamma-glutamyl ligase	2	4835.00	813.02	1591.90	3183.27	5180.26	6465.54	6.37
YP_887337.1	hypothetical protein MSMEG_3019	10	1525.00	324.33	335.07	971.10	2033.73	2071.53	6.27
YP_888583.1	hypothetical protein MSMEG_4306	3	1240.00	262.27	108.23	823.69	1628.69	1633.49	6.21
YP_888841.1	hypothetical protein MSMEG_4572	6	3300.00	854.98	916.81	1689.76	5158.94	6004.69	6.03
YP_889750.1	serine/threonine-protein kinase PknE	17	1750.00	495.68	340.03	1335.96	2779.86	1971.14	5.61
YP_888931.1	oxidoreductase alpha (molybdopterin) subunit	14	6435.00	1958.25	1772.28	3814.39	10104.71	9787.28	5.16
YP_885585.1	TROVE domain-containing protein	1	5350.00	4502.22	5330.38	13175.32	22938.11	45172.06	5.10
YP_889314.1	malyl-CoA lyase	9	3170.00	710.74	1208.58	2745.63	3427.93	3257.86	4.82
YP_887229.1	transcriptional regulator, GntR family protein	3	23400.00	5498.83	3815.15	14557.35	26348.99	23544.23	4.79
YP_888942.1	hypothetical protein MSMEG_4681	9	934.00	686.26	827.07	620.99	3146.53	3120.98	4.59
YP_885131.1	IS1137, transposase orfB	1	589.00	267.51	471.71	669.82	1219.39	1733.88	4.56
YP_885853.1	50S ribosomal protein L18	5	35700.00	8672.24	4617.86	24784.34	39310.30	19346.42	4.53
YP_890290.1	ribosomal protein L28	3	740.00	532.40	397.75	590.59	2396.14	1487.48	4.50
YP_889219.1	kinase	1	10700.00	1940.76	1307.79	13728.13	8697.72	6944.63	4.48
YP_887797.1	putative secreted protein	2	10800.00	3033.54	2205.21	5670.92	12833.41	12912.82	4.23
YP_888186.1	hypothetical protein MSMEG_3896	4	426.50	141.62	176.33	330.30	593.49	513.32	4.19
YP_891115.1	hypothetical protein MSMEG_6920	3	266.00	150.37	245.32	384.20	618.22	380.54	4.11
YP_889745.1	hypothetical protein MSMEG_5507	2	120.00	228.61	204.29	138.20	935.86	649.29	4.09
YP_890272.1	cobalamin synthesis protein/P47K	2	750.00	6176.35	2435.20	1557.08	24814.09	11954.63	4.02
YP_886110.1	transmembrane protein	3	4530.00	6469.21	8586.33	5002.94	25837.36	37962.78	3.99
YP_885379.1	hypothetical protein MSMEG_0977	3	19100.00	7404.62	5492.73	17782.07	29418.77	26008.16	3.97
YP_889430.1	alpha-methylacyl-CoA racemase, putative	8	3830.00	1232.65	1177.01	2206.64	4813.59	5630.54	3.91
YP_888359.1	butyryl-CoA dehydrogenase	1	30200.00	6032.10	5862.52	52332.73	23023.39	17795.06	3.82
YP_888187.1	proteasome component	9	8160.00	2535.23	3283.01	6974.63	9550.44	10768.29	3.77
YP_888936.1	ATP-dependent Clp protease proteolytic subunit	1	216.00	104.91	108.23	110.56	385.43	661.61	3.67
YP_887771.1	AMP-binding enzyme	5	1410.00	501.80	548.37	1253.04	1790.71	1861.64	3.57
YP_887617.1	universal stress protein family protein, putative	4	5450.00	1320.07	1366.42	3662.37	4664.37	4302.75	3.53
YP_885332.1	ErfK/YbiS/YcfS/YnhG family protein	10	497.00	739.15	900.12	436.72	2498.46	1451.44	3.38
YP_885700.1	TnpC protein	1	5970.00	3365.74	3896.32	5878.22	11341.15	25916.90	3.37
YP_886388.1	short chain dehydrogenase	8	3025.00	974.75	1402.49	1939.44	3244.59	9335.56	3.33
YP_889845.1	spore protein	2	7605.00	11233.70	10642.72	8264.52	37349.05	47909.76	3.33
YP_890530.1	cytochrome P450 107B1	3	1680.00	1932.02	2868.12	2856.19	6335.69	7108.90	3.28
YP_889680.1	methionyl-tRNA synthetase	6	8650.00	2548.34	2340.50	6767.32	8177.57	8454.93	3.21
YP_890261.1	2-hydroxy-6-ketonona-2,4-dienedioic acid hydrolase	6	2430.00	1028.95	1609.94	1575.51	3261.65	3107.29	3.17
YP_889159.1	rhomboid family protein	3	2420.00	789.42	705.31	2340.23	2481.41	1359.72	3.14
YP_889661.1	hypothetical protein MSMEG_5421	4	287.50	440.17	635.86	276.41	1325.98	3034.29	3.01
YP_887804.1	hypothetical protein MSMEG_3500	8	4600.00	2098.12	2069.92	4063.16	6288.80	6278.46	3.00

NCBI ID	Protein	# pep.	TMT126 (wt 0h)	TMT127 (wt 5h)	TMT128 (wt 11h)	TMT129 (clpP2 0h)	TMT130 (clpP2 5h)	TMT131 (clpP2 11h)	clpP2/wt 5h
YP_887416.1	glucose-6-phosphate 1-dehydrogenase	13	1380.00	497.43	611.51	667.98	1475.20	1679.12	2.97
YP_890470.1	hypothetical protein MSMEG_6251	3	4290.00	1862.08	1821.89	3703.83	5508.56	6415.35	2.96
YP_889537.1	acyl-CoA synthetase	1	4040.00	1477.43	1533.27	4081.58	4357.39	3203.11	2.95
YP_887239.1	aldo/keto reductase	1	120.00	1626.04	1614.45	2948.32	4732.58	4946.11	2.91
YP_886323.1	hypothetical protein MSMEG_1957	18	894.50	750.08	942.51	740.31	2178.69	2710.32	2.91
YP_884783.1	hypothetical protein MSMEG_0370	17	5670.00	2640.14	1794.83	5620.24	7606.24	7893.70	2.88
YP_890805.1	ATP-binding protein	4	2025.00	846.68	1100.35	1796.63	2421.72	2920.21	2.86
YP_886528.1	superfamily protein I DNA or RNA helicase	4	11720.00	3552.38	3130.13	8889.65	10128.16	9737.55	2.85
YP_889373.1	hypothetical protein MSMEG_5127	6	677.00	583.54	662.46	863.31	1654.27	2345.30	2.84
YP_890257.1	hypothetical protein MSMEG_6033	6	703.00	818.27	948.37	678.57	2319.39	1874.41	2.84
YP_886128.1	endonuclease VIII and DNA n-glycosylase with an AP lyase activity	1	2430.00	1835.86	2281.87	4081.58	5201.58	6214.58	2.83
YP_889001.1	clavaldehyde dehydrogenase	2	10790.00	3151.55	4496.10	6822.60	8761.68	9933.29	2.78
YP_884446.1	putative septation inhibitor protein	3	2300.00	1162.71	1397.98	2128.32	3231.80	2874.59	2.78
YP_890299.1	transcriptional regulator, CarD family protein	40	1350.00	697.63	701.25	990.45	1858.93	2071.53	2.67
YP_888640.1	hypothetical protein MSMEG_4365	2	676.50	576.98	650.74	772.09	1526.37	2111.68	2.65
YP_888514.1	transmembrane protein	2	68250.00	79291.54	87576.97	53208.02	209086.45	228232.99	2.64
YP_888491.1	glutamine amidotransferase	3	15700.00	5254.05	4392.38	12069.70	13814.03	11498.34	2.63
YP_885452.1	hexapeptide transferase family protein	18	2755.00	1792.15	1988.75	2717.99	4707.00	5516.47	2.63
YP_886165.1	dehydrogenase	1	588.00	143.37	427.51	592.43	362.41	280.16	2.53
YP_889350.1	indolepyruvate ferredoxin oxidoreductase	10	2565.00	1897.05	1411.51	3381.36	4792.28	5165.13	2.53
YP_885199.1	hypothetical protein MSMEG_0794	1	7370.00	4965.56	8892.99	5242.49	12364.41	12228.40	2.49
YP_887869.1	FadD16 protein	4	4710.00	3330.77	3887.30	4551.47	8279.89	8833.65	2.49
YP_888615.1	NAD/mycothiol-dependent formaldehyde dehydrogenase	50	3300.00	1687.24	1803.85	2197.42	4169.79	5940.81	2.47
YP_886244.1	S30AE family protein	1	9300.00	10315.77	15603.31	9130.59	25325.72	30844.76	2.46
YP_885675.1	TPR repeat-containing protein	3	1360.00	591.85	928.98	1234.61	1449.62	1551.36	2.45
YP_889261.1	hypothetical protein MSMEG_5010	2	2273.50	2304.44	2523.14	2727.66	5614.30	7446.55	2.44
YP_884959.1	ISMsm5, transposase	10	1730.00	1289.47	1065.17	1833.49	3138.00	4695.16	2.43
YP_885614.1	ISMsm6, transposase	2	2347.50	2516.87	1933.28	2551.68	6101.20	7711.19	2.42
YP_884473.1	hypothetical protein MSMEG_0055	2	3055.00	1276.36	1781.30	3367.54	3086.84	4800.10	2.42
YP_884796.1	hypothetical protein MSMEG_0383	10	644.00	477.76	1470.14	395.26	1151.17	650.66	2.41
YP_887260.1	NAD dependent epimerase/dehydratase family protein	3	12700.00	5629.96	5844.48	12253.97	13472.94	9764.47	2.39
YP_887795.1	hypothetical protein MSMEG_3491	6	3495.00	1870.83	4153.37	2962.14	4446.92	4676.91	2.38
YP_891092.1	replicative DNA helicase	8	5955.00	2723.19	4951.57	4924.62	6403.91	15180.55	2.35
YP_887087.1	hypothetical protein MSMEG_2756	7	941.00	771.06	983.10	1041.13	1807.76	2619.07	2.35
YP_886830.1	D-lactate dehydrogenase	1	212.00	104.91	108.23	181.51	245.58	204.41	2.34
YP_886345.1	hypothetical protein MSMEG_1981	3	120.00	219.43	108.23	110.56	511.63	377.80	2.33
YP_885429.1	monooxygenase	31	1030.00	858.48	1470.14	1151.69	1995.36	2619.07	2.32
YP_888563.1	lipoyl synthase	4	10950.00	6976.26	9064.35	9766.32	16201.64	15650.52	2.32
YP_889372.1	FO synthase	4	1820.00	1128.61	1350.63	1907.20	2596.53	3741.52	2.30
YP_888098.1	hypothetical protein MSMEG_3806	6	5155.00	3304.54	3476.92	4486.98	7563.61	9413.13	2.29
YP_886850.1	lactate 2-monooxygenase	2	3095.00	1599.82	761.23	2644.28	3649.63	4101.99	2.28
YP_885727.1	ribosomal protein L33	40	5720.00	2207.40	1591.90	5002.94	5018.25	4772.73	2.27
YP_888037.1	GTP-binding protein EngA	8	2605.00	1175.82	1170.70	2335.62	2647.69	2568.88	2.25
YP_885528.1	heat shock protein HtpX	1	571.00	581.35	901.93	589.66	1304.66	1806.88	2.24
YP_890529.1	saccharopine dehydrogenase	3	8130.00	7990.35	8938.08	9130.59	17907.08	31757.33	2.24

NCBI ID	Protein	# pep.	TMT126 (wt 0h)	TMT127 (wt 5h)	TMT128 (wt 11h)	TMT129 (clpP2 0h)	TMT130 (clpP2 5h)	TMT131 (clpP2 11h)	clpP2/wt 5h
YP_887529.1	ABC-type molybdenum transport system, ATPase component/photorepair protein PhrA	3	21300.00	10927.72	21014.87	55281.06	24387.73	15787.41	2.23
YP_886211.1	rubredoxin	1	12900.00	3374.48	10281.95	13451.72	7478.34	3924.04	2.22
YP_888108.1	excinuclease ABC subunit B	11	2060.00	2474.04	2236.78	2423.15	5457.40	6962.89	2.21
YP_887936.1	inositol-5-monophosphate dehydrogenase	16	193.50	104.91	120.86	166.76	230.23	320.77	2.20
YP_885030.1	PPE family protein	2	120.00	104.91	108.23	150.64	228.53	136.89	2.18
YP_890280.1	cadmium transporting P-type ATPase	2	1880.00	4365.84	2768.91	1805.85	9332.15	5393.27	2.14
YP_891101.1	hypothetical protein MSMEG_6901	2	7530.00	7015.60	7598.72	7614.97	14752.02	21673.46	2.10
YP_884953.1	ANTAR domain-containing protein	6	9960.00	5940.31	6272.89	11254.30	12108.60	10590.34	2.04
YP_889351.1	ABC transporter ATP-binding protein	2	6215.00	5485.72	5303.32	9591.26	11127.97	16234.57	2.03
YP_885118.1	chaperone protein DnaJ	17	1570.00	2194.29	2750.87	2155.96	4451.19	5329.39	2.03
YP_888525.1	1-acylglycerol-3-phosphate O-acyltransferase	3	3360.00	2954.86	3120.66	3565.63	5986.08	8559.88	2.03
YP_887058.1	recombination regulator RecX	3	6600.00	3619.26	2471.28	4920.01	7333.37	6634.36	2.03
YP_890829.1	hypothetical protein MSMEG_6620	1	5880.00	5061.72	7287.56	6440.24	10232.62	17247.52	2.02
YP_891139.1	chromosomal replication initiation protein	4	5455.00	2570.20	2254.81	2925.29	5184.53	5730.92	2.02
YP_887414.1	6-phosphogluconolactonase	6	2485.00	1311.33	1379.95	2413.94	2639.16	3499.69	2.01
YP_890515.1	aldehyde dehydrogenase	1	28400.00	15910.76	13077.92	54083.30	31976.93	17247.52	2.01

UNDER-REPRESENTED PROTEINS

NCBI ID	Protein	# pep.	TMT126 (wt 0h)	TMT127 (wt 5h)	TMT128 (wt 11h)	TMT129 (clpP2 0h)	TMT130 (clpP2 5h)	TMT131 (clpP2 11h)	clpP2/wt 5h
YP_888935.1	ATP-dependent Clp protease proteoly	20	12950	13768.927	13619.076	11655.09	7559.3451	9764.46629	0.549
YP_890980.1	hypothetical protein MSMEG_6772	8	13590	12671.784	14611.194	14373.075	6233.3686	15253.5565	0.492
YP_886029.1	O-acetylhomoserine aminocarboxypro	2	486	579.16915	405.41553	350.11337	281.8233	1559.57691	0.487
YP_884500.1	phosphocarrier protein HPr	3	3260	5525.0552	7882.8292	3298.4365	2651.953	3139.23028	0.480
YP_888642.1	formamidase	2	13125	12357.066	12572.842	11733.405	5909.3357	10165.9957	0.478
YP_885552.1	5-oxovalerate dehydrogenase	2	2138	2266.4091	2611.9764	2146.7478	1076.5565	1266.64292	0.475
YP_886058.1	flavin-containing monooxygenase FM	2	113700	103813.34	50417.638	115583.48	47922.752	8692.20013	0.462
YP_888190.1	diaminobutyrate--2-oxoglutarate amin	4	3905	7269.1193	8198.5031	4076.978	3334.1274	2459.36791	0.459
YP_890773.1	nonspecific lipid-transfer protein	3	4300	4519.7049	4139.8382	4505.4062	2012.4145	1496.60979	0.445
YP_884970.1	hypothetical protein MSMEG_0559	7	8490	16784.978	6926.7881	7997.3264	7017.8691	8176.59981	0.418
YP_887577.1	maltose/maltodextrin-binding protein	3	566	840.1231	690.87496	448.69792	338.52905	158.786648	0.403
YP_888096.1	lipoprotein LpqH	5	630	2307.9344	4617.8588	770.24941	895.3539	1022.07498	0.388
YP_884657.1	hypothetical protein MSMEG_0242	1	4080	8602.3011	11093.684	3851.247	3300.0187	3111.85328	0.384
YP_886054.1	aspartate ammonia-lyase	1	14200	12938.42	8369.869	15755.102	4928.71	3285.241	0.381
YP_889726.1	DNA-binding response regulator	1	340	274.50432	195.71784	372.22579	102.32616	109.508033	0.373
YP_885732.1	transcription antitermination protein N	5	1280	1818.3726	454.57047	1124.0482	410.15736	397.879187	0.226

Appendix 4 – Quantitative PCR to assess regulation of the putative WhiB1 regulon in Msm upon expression of GFP-WhiB1 and GFP-WhIB1

Quantitative PCR, using probe sets that hybridize to essential genes in the putative WhiB1 regulon to determine transcriptional modulation in wildtype Msm, or Msm inducibly expressing GFP-WhiB1 or WhiB1-GFP on tetracycline (ATc) inducible, episomal plasmids. The putative regulon in Msm was determined by finding homologs for Mtb genes identified as potential downstream targets of WhiB1 through ChIP-Seq data deposited in the Tuberculosis Database (www.tbdb.org).

RNA was isolated from cultures grown for 6 hours from a starting OD_{600} of 0.06 in the presence of the inducer ATc (100 ng/mL). Relative standard curves were generated for each probe set, and sigA transcript was used as an endogenous control. Data are represented as mean fold change, normalized to transcript in wildtype cultures +/- SEM of technical replicates.